



2017

Essays On Applied Microeconomics: Airline Networks And Job Search

Kristijan Gjorgjevik

University of Pennsylvania, kristijan.gjorgjevik@gmail.com

Follow this and additional works at: <https://repository.upenn.edu/edissertations>



Part of the [Labor Economics Commons](#)

Recommended Citation

Gjorgjevik, Kristijan, "Essays On Applied Microeconomics: Airline Networks And Job Search" (2017). *Publicly Accessible Penn Dissertations*. 2307.

<https://repository.upenn.edu/edissertations/2307>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/2307>

For more information, please contact repository@pobox.upenn.edu.

Essays On Applied Microeconomics: Airline Networks And Job Search

Abstract

This dissertation consists of three chapters that develop and implement economic models to analyze modern problems in Industrial Organization and Labor Supply. In the first chapter, I extend the standard BLP model (Berry et al. [1995]) to account for capacity constraints in a network and evaluate the welfare effects of the 2013 merger between American Airlines and US Airways. I show that including capacity as a constraint in the profit maximization problem that airlines face generates better out-of-sample predictions and leads to different policy implications. In particular, I find that the merger increased consumer surplus by 1.5-1.7%, while the benchmark model could predict it to have decreased by as much as 4.5%. In other words, ignoring capacity constraints could lead regulators to erroneously believe that this merger harmed consumers. I find that, on average, the merger increased the variable profit margins of airlines by 0.3-0.4%, and American Airlines' by 2.5%. I develop and implement an approach for ex-post merger evaluation that could be useful in antitrust legislation.

In the second chapter, I extend the theory of efficiency wages (Shapiro and Stiglitz [1984]) to incorporate employer-sponsored health insurance. I develop sufficient conditions under which the Affordable Care Act increases efficiency wages. In particular, if the Affordable Care Act succeeds, at least in part, in inducing employers to provide health insurance and individuals to self-insure, then wages will rise after its implementation. I suggest that the Affordable Care Act may provide efficiency wage subsidies towards the welfare-maximizing wage level, and numerically show the existence of regions where this is the case.

In the third chapter, I extend Mirrlees' theory of optimal taxation (Mirrlees [1971]) to include endogenous job search. I use a public-use microdata file on the Canadian labor force to calculate the optimal, revenue-neutral federal tax rates. The results are highly sensitive to the level of inequality aversion chosen for the social welfare function. The optimal tax rate schedule is hump-shaped in income.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Economics

First Advisor

Katja Seim

Keywords

Airline Industry, Efficiency Wages, Industrial Organization, Labor Supply, Mergers, Mirrlees Taxation

Subject Categories

Economics | Labor Economics

ESSAYS ON APPLIED MICROECONOMICS: AIRLINE NETWORKS AND
JOB SEARCH

Kristijan Gjorgjevik

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

Katja Seim
Associate Professor of Business Eco-
nomics and Public Policy

Graduate Group Chairperson

Jesus Fernandez-Villaverde, Professor of Economics

Dissertation Committee
Petra Todd, Alfred L. Cass Term Professor of Economics
Kenneth Burdett, James Joo-Jin Kim Professor of Economics

ESSAYS ON APPLIED MICROECONOMICS: AIRLINE NETWORKS AND JOB
SEARCH

© Copyright

2017

Kristijan Gjorgjevik

This work is licensed under the

Creative Commons

Attribution-NonCommercial-ShareAlike 4.0

International License

To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/4.0/>

To my parents. Thank you for everything.

ABSTRACT

ESSAYS ON APPLIED MICROECONOMICS: AIRLINE NETWORKS AND JOB
SEARCH

Kristijan Gjorgjevik

Katja Seim

This dissertation consists of three chapters that develop and implement economic models to analyze modern problems in Industrial Organization and Labor Supply. In the first chapter, I extend the standard BLP model (Berry et al. [1995]) to account for capacity constraints in a network and evaluate the welfare effects of the 2013 merger between American Airlines and US Airways. I show that including capacity as a constraint in the profit maximization problem that airlines face generates better out-of-sample predictions and leads to different policy implications. In particular, I find that the merger increased consumer surplus by 1.5-1.7%, while the benchmark model could predict it to have decreased by as much as 4.5%. In other words, ignoring capacity constraints could lead regulators to erroneously believe that this merger harmed consumers. I find that, on average, the merger increased the variable profit margins of airlines by 0.3-0.4%, and American Airlines' by 2.5%. I develop and implement an approach for ex-post merger evaluation that could be useful in antitrust legislation.

In the second chapter, I extend the theory of efficiency wages (Shapiro and Stiglitz [1984]) to incorporate employer-sponsored health insurance. I develop sufficient conditions under which the Affordable Care Act increases efficiency wages. In particular, if the

Affordable Care Act succeeds, at least in part, in inducing employers to provide health insurance and individuals to self-insure, then wages will rise after its implementation. I suggest that the Affordable Care Act may provide efficiency wage subsidies towards the welfare-maximizing wage level, and numerically show the existence of regions where this is the case.

In the third chapter, I extend Mirrlees' theory of optimal taxation (Mirrlees [1971]) to include endogenous job search. I use a public-use microdata file on the Canadian labor force to calculate the optimal, revenue-neutral federal tax rates. The results are highly sensitive to the level of inequality aversion chosen for the social welfare function. The optimal tax rate schedule is hump-shaped in income.

Contents

List of Tables	ix
----------------	----

List of Figures	xii
-----------------	-----

1	An Empirical Analysis of the Airline Industry with Capacity Constraints	1
1.1	Introduction	1
1.2	Model	7
1.2.1	Demand	7
1.2.2	Supply	15
1.3	Data	19
1.3.1	DB1B	19
1.3.2	T-100	20
1.3.3	Evidence for Load Factors	21
1.3.4	Sample Selection	24
1.3.5	Adjustments	30
1.3.6	Airfare Taxes	33
1.3.7	Instruments	36
1.3.8	Identification	39

1.3.9	Empirical Capacity Constraints	39
1.4	Results	42
1.4.1	Demand	42
1.4.2	Supply	50
1.4.3	Prediction	57
1.4.4	Model Fit	59
1.4.5	Counterfactual: But-For AA-US Merger	63
1.5	Conclusion	91
2	Efficiency Wages with Health Insurance	94
2.1	Introduction	94
2.2	Benchmark Model (Pre-ACA)	99
2.2.1	Worker's Problem	99
2.2.2	Health Insurance Purchase Decision	104
2.2.3	Employer's Problem	106
2.2.4	Solving for No-Shirk-Conditions	108
2.2.5	Solving for Equilibrium Wage	114
2.3	Affordable Care Act	116
2.3.1	Worker's Problem	117
2.3.2	Employer's Problem	120
2.3.3	Insurance Exchange	121
2.3.4	Solving for No-Shirk-Conditions	125
2.4	Benchmark/ACA Comparison	129

2.4.1	Case 1: Firm offers HI after ACA	129
2.4.2	Case 2: Firm does not offer HI	133
2.5	Conclusion	142
3	Mirrlees Taxation with Endogenous Search Effort	144
3.1	Introduction	144
3.2	Model	145
3.2.1	Worker's Problem	146
3.2.2	Government's Problem	149
3.3	Data	150
3.4	Estimation	152
3.4.1	Limited Information MLE	152
3.4.2	Identification	155
3.4.3	Results	156
3.5	Conclusion	159
	Bibliography	161
	Appendices	165
A	Appendix to Chapter 1	166
B	Appendix to Chapter 2	179
C	Appendix to Chapter 3	203

List of Tables

1.1	Number of Links Operated $> 90\%$ Full	23
1.2	Percentage of Links Operated $> 90\%$ Full	23
1.3	Summary Statistics for the Observables (2012)	27
1.4	Other Summary Statistics (2012)	28
1.5	Summary Statistics for the Ticketing Carriers (2012)	29
1.6	Fuel Cost per Revenue Passenger Mile (cents/mile)	31
1.7	Summary Statistics for Selected Instruments	38
1.8	Demand-Side Parameter Estimates (2012)	46
1.9	Demand-Side Parameter Estimates (2015)	47
1.10	Price Elasticity of Demand	48
1.11	Willingness to Pay (% of Average Fare in 2012)	48
1.12	Willingness to Pay (% of Average Fare in 2015)	49
1.13	Fuel Component of Costs (% in 2012)	49
1.14	Supply-Side Parameter Estimates (2012)	53
1.15	Supply-Side Parameter Estimates (2015)	54
1.16	Estimated CPM (cents/mile)	60
1.17	Estimated Variable Profit Margin (%)	60

1.18	Estimated Median Fares	62
1.19	Intersection of Products	67
1.20	Intersection of Products (Alternate)	67
1.21	Intersection of Routes	68
1.22	Intersection of Markets	69
1.23	Intersection of Origins	70
1.24	Visualization of Product Allocation (n=183761)	73
1.25	Products Remaining from 2012	74
1.26	Robustness Check: Merging US and AA in 2012	76
1.27	Counterfactual CPM (cents/mile)	78
1.28	Counterfactual Median Fares	79
1.29	Counterfactual Variable Profit Margin (%)	79
1.30	Counterfactual Summary of Merger: All Carriers	80
1.31	Counterfactual CPM (cents/mile), Quality Transfer	81
1.32	Counterfactual Median Fares, Quality Transfer	82
1.33	Counterfactual Variable Profit Margin (%), Quality Transfer	82
1.34	Counterfactual Summary of Merger: All Carriers, Quality Transfer	83
1.35	Counterfactual CPM (cents/mile), Full Model	85
1.36	Counterfactual Capacity Constrained Links, Full Model	85
1.37	Counterfactual Median Fares, Full Model	86
1.38	Counterfactual Variable Profit Margin (%), Full Model	87
1.39	Counterfactual Summary of Merger: All Carriers, Full Model	88
1.40	Counterfactual CPM (cents/mile), Full Model and Quality Transfer	89

1.41	Counterfactual Capacity Constrained Links, Full Model and Quality Transfer	89
1.42	Counterfactual Median Fares, Full Model and Quality Transfer	90
1.43	Counterfactual Variable Profit Margin (%), Full Model and Quality Transfer	90
1.44	Counterfactual Summary of Merger: All Carriers, Full Model and Quality Transfer	91
2.1	Some Possible Values of n_h^{EX}	124
3.1	Data Summary by Province	151
3.2	2014 Tax Rate Schedule - Canada	151
3.3	Optimal Tax Rates using Ontario Subsample	158
3.4	Optimal Tax Rates	159
A.1	Summary Statistics for the Observables (2015)	167
A.2	Other Summary Statistics (2015)	167
A.3	Summary Statistics for the Ticketing Carriers (2015)	168
A.4	Demand-Side Parameter Estimates (2015)	169
A.5	Supply-Side Parameter Estimates (2015)	170

List of Figures

1.1	Histogram of American Airlines' Load Factors on Links	22
1.2	Distribution of Fares, Non-Weighted vs. Weighted (2012)	28
1.3	Average Domestic Fuel Cost per Gallon	30
1.6	Distribution of Fuel Cost as a Fraction of Total Variable Cost	50
1.7	Reduction in Marginal Costs vs. κ	56
1.8	Negative Marginal Costs Fully Explained by Residuals	56
2.1	Job Search over the Lifecycle - USA (Source: Aguiar et al. [2013])	95
2.2	Job Search over the Lifecycle - OECD Countries (Source: Aguiar et al. [2013])	96
2.3	Supporting Data from CPS and MEPS	97
3.1	Provincial Tax Functions	152
3.2	Optimal Federal Tax Rates vs. Inequality Aversion (Ontario)	158
3.3	Optimal Federal Tax Rates vs. Inequality Aversion (All Provinces)	159
A.1	BIC and % Products with a Capacity Constrained Link vs. κ	172
B.1	Benchmark - No Health Considerations	181

B.2	MEPS Health Expenditure Data and Health Transitions	182
B.3	HI Offered: ACA Effects on HI Coverage while Unemployed	197
B.4	HI Offered: ACA Effects on Profits and Wage	197
B.5	Complete Decision: ACA Effects on Firm's Choice of HI offer	199
B.6	Complete Decision: ACA Effects on Profits and Wage	199
B.7	Complete Decision: ACA Effects on Self-Insurance Decision of Unemployed Workers	200
B.8	Complete Decision: ACA Effects on Self-Insurance Decision of Employed Workers	200
B.9	Wage as Subsidies Decrease, Plotted Against Constant Benchmark Wage .	201

1 An Empirical Analysis of the Airline Industry with Capacity Constraints

1.1 Introduction

Bankruptcies and mergers are common and costly to the United States airline industry. Since the turn of the century, all major legacy carriers have filed for, and emerged from, Chapter 11 bankruptcy protection at least once. US Airways (“US”) first filed for bankruptcy protection on August 11, 2002 and again on September 12, 2004; United Airlines (“United”) filed on December 9, 2002; Delta Air Lines (“Delta”) filed on September 14, 2005; and American Airlines (“AA”) filed on November 29, 2011.

Since the Great Recession, there have been 7 major mergers amongst airlines in the United States. Southwest acquired ATA Airlines (November 19, 2008) and AirTran Airways (September 27, 2010); Republic Airways acquired Midwest Airlines (June 23, 2009) and Frontier Airlines (August 14, 2009); Delta acquired Northwest Airlines (April 14, 2008); United acquired Continental Airlines (May 3, 2010); and AA acquired US (February 14, 2013).

In recent years, in addition to changes in the ownership of products, the United States airline industry changed markedly in another dimension as well: its load factors.

In particular, the number of flights that were operated at or close to capacity¹ markedly increased. For example, from the third quarter of 2012 (before the most recent airline merger between AA and US) and the third quarter of 2015 (after said merger), the number of flights that Southwest and AA operated at or close to capacity more than doubled.

Given the continually changing structure of the United States airline industry through bankruptcies and mergers, accurate cost and merger analyses are important. Given the large fixed costs of operating a flight, the changes in load factors can dramatically affect the marginal costs per passenger of operating a route. For example, the cost per passenger of flying a single passenger in a Boeing 747, averaged over different capacity levels, is dramatically higher than the cost per passenger of flying a Boeing 747 operated at capacity. As a result, any accurate cost and merger analyses need to take capacity considerations into account.

This paper builds on recent literature on estimating differentiated products with aggregate data in the spirit of Berry et al. [1995] (“BLP”). Notably, Berry et al. [2006] and, more recently, Berry and Jia [2010] have applied the techniques in BLP to provide estimates of a model of airline competition with differentiated products. But their approach has supply-side shortcomings: it assumes there are constant marginal costs per passenger and no network effects.

To understand what I mean by “network effects”,² consider what happens when an airline lowers the price on one of its direct connections. For example, consider United lowering the price of its ORD-JFK route. The BLP prediction is that consumers substitute

¹The threshold “at or close to capacity” will be defined as a load factor greater or equal to 90%.

²Barla and Constantatos [2000] and Pels [2008] provide theoretical results on the optimal network structure (hub and spoke vs. point to point).

towards the ORD-JFK route and away from similar flights (connecting or otherwise) from United and other carriers, as if these products were operated independently of one another. This is not the case. In particular, we might think that the increase in passengers towards the ORD-JFK route could have secondary effects on the utility of passengers (demand-side network effect), or on the ability of United to fulfill all the demand (supply-side network effect) — on this route *or on any routes that use this segment as a connecting flight*.

Consider the demand-side first: flow can affect utility. This is to say, the increasing number of passengers now flying ORD-JFK could affect each individual passenger’s utility from flying the ORD-JFK route, either through congestion (negative effect) or more scheduled flights (positive effect). Wei [2014] explicitly models this second effect and generates higher supply-side marginal cost estimates than the existing literature. He does this by including “positive utility of frequency” in the individual’s problem to capture network effects. He justifies this by claiming that individuals care about frequency in that it reduces the cost of delays — one can more quickly catch the next flight.

On the supply-side, consider the obvious constraint: capacity. Flow cannot increase indefinitely; indeed, each route has some capacity (the number of seats on scheduled flights), and changing that capacity is not feasible in the short-run.³ In other words, the firm does not face an unconstrained optimization problem, and capacity on a single link can affect a firm’s pricing on its entire network. The role of this paper is to provide a theoretical foundation for constrained multiproduct optimization problems in a network,

³See e.g. Barnhart et al. [2003]. Roughly speaking, an airline scheduling process includes four steps. The airline (i) first allocates to each city-pair planes by type (e.g. two 737 flights a day between Chicago and Cleveland), then (ii) allocates specific planes (i.e., tail numbers), then (iii) determines a maintenance schedule, and (iv) assigns crews. The procedure is done using demand forecasts and flights are typically scheduled more than six months before takeoff.

and to estimate a model with an empirical notion of capacity constraints in the observed network of airline products. I show that the inclusion of capacity constraints is important in accurately predicting realized costs, prices, and profits of airlines; ignoring these, as previous literature has, leads to inaccurate predictions.

While the pricing decisions in this chapter are static, there is a large literature on dynamic pricing. In particular, Lazarev [2013] studies how a firm’s ability to price discriminate over time affects production, product quality, and product allocation among consumers. Williams [2013] goes one step further and takes into account the effects of stochastic demand in a model with limited capacity. Due to the complexity of the problem, these authors must restrict themselves to the considerations of monopoly routes only. Moreover, such an analysis requires high-frequency (daily) data, which is not publically available. I leave dynamic pricing considerations for future work.

There is also a growing literature on estimating static models of competition while allowing for market structure to be endogenous. Some works include Reiss and Spiller [1989], Ho [2009], Draganska et al. [2009], and Benkard et al. [2010]. More recently, Roberts and Sweeting [2014] estimate a model of endogenous entry for the airline industry but only consider *sequential* move equilibria. On the other hand, Ciliberto et al. [2016] consider the *simultaneous* entry and pricing problem. They find that post-merger entry mitigates the price effects of a merger, and the merged firm has a strong incentive to enter new markets. While I take the network structure as exogenous in this chapter, I perform different robustness checks on the fixed network structure using pre-merger data, post-merger data, or a combination of both. The results in this chapter are robust to a host of resulting post-merger network structures.

Due to the recency of the AA-US merger, there are few papers that explore its welfare effects. Indeed, while Ciliberto et al. [2016] perform a counterfactual exercise on the AA-US merger, they focus on the route entry problem, and do not perform a welfare analysis. Rupp and Tan [2016] explore the effects of 5 mergers between 1998 and 2015, and find that airlines offer improved product quality at de-hubbed airports after a merger, due to more reliable flight schedules and shorter travel times. The nature of their reduced-form model does not allow them to perform a welfare analysis. Kwoka and Shumilkina [2010] explore the effects of an earlier merger between US-Piedmont (in 1989), and find that prices rose by 5-6% on routes that one carrier served and the other was a potential entrant. However, Luo [2014] finds that the fares for airport-pairs where Delta and Northwest competed with each other prior to their 2008 merger did not increase. I find that average fares increased by 0.6-1.2% because of the merger, and go one step further by performing a welfare analysis, highlighting the effects of capacity constraints.

I find that the merger increased consumer surplus by 1.5-1.7%. Ignoring capacity constraints and economies of scale, the same analysis could have found a decrease in consumer surplus as high as 4.5%. The difference between these numbers is not obvious: there are two competing forces introduced by capacity constraints and economies of scale. On the one hand, due to economies of scale, airlines want to keep flights operating at or close to capacity, so they lower prices of particular routes to drive up demand. This lowers marginal costs, and allows them to price even lower, increasing consumer surplus. On the other hand, due to capacity constraints, airlines cannot increase demand indefinitely, so they are forced to limit the demand of certain routes, especially those that are used by many connecting flights. This leads them to price higher, decreasing consumer surplus.

In short, this chapter contributes to the large literature on the estimation of models of airline competition with differentiated products using aggregate data by introducing capacity-constraint considerations on the supply-side,⁴ the estimates of which are shown to be superior to benchmark models in prediction. I find that, in 2012, flights operated at or near capacity had lower marginal costs per passenger by \$128. In 2015, this discount reduced to \$57. I also develop a robust approach for ex-post merger evaluation that can be used for future antitrust work. I use it to simulate a counterfactual but-for world without the AA-US merger, and perform a welfare analysis. I find that the AA-US merger benefited consumers and airlines — specifically AA, which enjoyed 2.5% higher profit margins directly as a result of the merger.

This chapter has contributions above and beyond a pure “case study” of the AA-US merger. In particular, it provides a framework for estimating general equilibrium models using aggregate product-level data in a network with constraints. In this sense, we can think of the airline industry as a good example of an observable network with observable constraints (capacity), and the AA-US merger as a means to test the theory developed in this paper and show that network considerations can drastically affect policy implications. In this case, I find that benchmark models would predict that the merger hurts consumer welfare, when in fact it is welfare-enhancing. The theory and approach developed in this paper could affect future policy decisions both within the airline industry, and in other industries with network constraints.

A limitation of this approach is that the full structural model cannot be estimated:

⁴There are also contributions on the demand-side: I am able to construct a previously unidentified “stopover” variable by comparing two datasets: one that does not distinguish between direct and nonstop flights, and another that includes only nonstop flights. I find that stopovers require 1/3 to 1/5 of the fare discount that a layover does.

capacity constraints never bind empirically. For this reason, a reduced-form model is used for estimation, while the full model is used for the counterfactuals. Moreover, this paper abstracts from strategic capacity considerations: network structure is assumed to be fixed.

The paper is organized as follows. Section 2 outlines the economic model. Section 3 discusses the airline data, sample selection, instruments, and accounting adjustments. Section 4 presents the results, and demonstrates the usefulness of the results for prediction: first in establishing model fit, and then in performing counterfactual exercises. Section 5 concludes.

1.2 Model

1.2.1 Demand

I define a product, j , as any distinct way of getting from point A to point B and back, in a given fare bin, *with a given set of operating carriers*. This is to say, a direct flight from Philadelphia to London purchased with US, a connecting flight via New York purchased with US, and a connecting flight via New York purchased with United are all distinct products, j . I use the term “purchased” intentionally; I differentiate products according to their ticketing carrier. Moreover, a connecting flight via New York with the Philadelphia-New York link operated by Delta Connection is a different product than, *ceteris paribus*, that link operated by United Express. Finally, a direct flight with United priced at \$300 is a different product than the same flight priced at \$400.

While it may seem odd to define products conditional on a strategic variable (price),

this is standard and necessary in the airline literature (see Berry and Jia [2010] for a more elaborate discussion). In short, the necessity is driven by data limitations; there are large price variations in seemingly identical products (but-for the price) in the data, because many product characteristics are unobserved. For example, I do not know the class of travel (Economy, Economy Plus, Business, etc.). I also do not know the time or day of departure (red-eye, daytime, weekend, holiday, etc.).⁵ There are further unobserved product characteristics, such as whether a ticket was purchased with frequent flyer miles,⁶ whether there was advance-purchase, or whether there were length-of-stay requirements, etc. For this reason, it is particularly important to allow for product-unobservable characteristics that are correlated with price.

The suffix “*with a given set of operating carriers*” makes this product definition slightly different than that of Berry and Jia [2010]. Since a goal of this paper is to determine which products have capacity-constrained links, it is important to differentiate products according to the operating carrier on *each* of their links. For example, while previous literature would have treated a flight ticketed by United but operated by American as identical to a flight ticketed and operated by United, these two flights could have different load factors. As a consequence, they must be distinct products on the supply-side. By extension, they must be distinct products on the demand-side as well. This leads to a larger set of products than in Berry and Jia [2010].

The demand model is a random-coefficient discrete-choice model as in McFadden [1973] and BLP. As in Berry and Jia [2010], I use a discrete-type version of the random

⁵All I know is the year and quarter of departure, though I abstract from seasonality by using only one quarter of data in estimation.

⁶Though these are mostly filtered out by excluding abnormally low fares.

coefficient model, with two types namely: tourists and business travelers. I define a market, t , as any *directional* pair of cities, $A \rightarrow B$. In market t , the utility of consumer i who is of type r and purchases product j is given by:

$$u_{ijt} = x_{jt}\beta_r - \alpha_r p_{jt} + \xi_{jt} + v_{it}(\lambda) + \lambda \epsilon_{ijt} \quad (1.2.1)$$

where

1. x_{jt} is a vector of product characteristics
2. β_r is a vector of “tastes for characteristics” for consumers of type r
3. α_r is the marginal disutility of a price increase for consumers of type r
4. p_{jt} is the product price
5. ξ_{jt} is the unobserved (to researchers) characteristic of product j
6. v_{it} is a “nested logit” random taste that is constant across airline products and differentiates “air travel” from the “outside good”
7. $\lambda \in [0, 1]$ is the nested logit parameter
8. ϵ_{ijt} is an iid (across products and consumers) “logit error”

where the error structure, $v_{it}(\lambda) + \lambda \epsilon_{ijt}$, is assumed to be generalized extreme value.

Let the utility of the outside good (think of this as driving, taking the train, or making a phone call) be given by:

$$u_{iot} = \epsilon_{iot} \quad (1.2.2)$$

Using random coefficients introduces a correlation in utility between products with similar characteristics, but I also want to introduce a correlation in utility between products in the same market. The reason for this is that I believe the outside good in each market is not at all similar to the “inside” goods (the airlines’ products in a market). To do this, I impose two nests *within each market*: one with all the airline products in a market, and the other with just the outside option of not flying. This choice of nests allows for correlations in the unobservables of different airline products, while maintaining independence from the outside option of alternate transportation methods. I estimate a single set of parameters across all markets.⁷

The share of type r consumers who purchase any airline product in market t , s_t^r , is given by:

$$s_t^r = \frac{D_{rt}^\lambda}{1 + D_{rt}^\lambda} \quad (1.2.3)$$

where

$$D_{rt} = \sum_{k=1}^{J_t} e^{(x_{kt}\beta_r - \alpha_r p_{kt} + \xi_{kt})/\lambda} \quad (1.2.4)$$

where J_t is the number of products in market t .

Then, the share of type r consumers that purchase product j in market t , and the share

⁷There are a few approaches to capture consumer differences across markets. In order of estimating most to least parameters, one could: include market fixed effects, allow parameters to vary across regions (Northeast, Southwest, etc.), or include region fixed effects. Alternatively, to just capture the effect of income differences across markets, one could aggregate the income per capita in the endpoint cities of each market and estimate a single parameter.

of type r consumers that do not purchase a flight in market t are given by, respectively:

$$s_{jt}^r = \frac{D_{rt}^\lambda}{1 + D_{rt}^\lambda} \frac{e^{(x_{jt}\beta_r - \alpha_r p_{jt} + \xi_{jt})/\lambda}}{D_{rt}} \quad (1.2.5)$$

$$s_{0t}^r = \frac{1}{1 + D_{rt}^\lambda} \quad (1.2.6)$$

Let γ_r denote the percentage of type r consumers in the population. Then, the overall market share of product j in market t is given by:

$$s_{jt} = \sum_r \gamma_r s_{jt}^r \quad (1.2.7)$$

Let M_t denote the effective size of the market. Empirically, I will estimate this as μG_t , where G_t is the geometric mean of the Metropolitan Statistical Area (“MSA”) populations of the endpoint cities of the market, and μ is a parameter to be estimated. The use of the geometric mean has both empirical and (weak) theoretical precedents in the literature on travel demand. For a more elaborate discussion, see Berry et al. [2006]. Then, the overall market demand of product j in market t is given by:

$$q_{jt} = M_t s_{jt} \quad (1.2.8)$$

A central assumption in this paper is that network structure is exogenous. In particular, I take route composition and fleet allocation as fixed. Amongst other things, this allows me treat the conventionally endogenous variable “number of direct flights” as exogenous, and include it on the demand side without instrumenting for it. For the rest of this section, when there is no confusion, I may omit some subscripts on these variables.

Regression, 2SLS

Using the product share equations from above, I get that:

$$\begin{aligned}\ln s_j^r &= (x_j \beta_r - \alpha_r p_j + \xi_j) / \lambda + \ln \left(\frac{D_r^{\lambda-1}}{1 + D_r^\lambda} \right) \\ \ln s_0^r &= \ln \left(\frac{1}{1 + D_r^\lambda} \right) \\ \ln s_j^r - \ln s_0^r &= (x_j \beta_r - \alpha_r p_j + \xi_j) / \lambda + (\lambda - 1) \ln D_r\end{aligned}$$

Moreover,

$$\begin{aligned}\ln s_t^r &= \ln \left(\frac{D_r^\lambda}{1 + D_r^\lambda} \right) \\ \Rightarrow \ln D_r &= (\ln s_t^r - \ln s_0^r) / \lambda\end{aligned}$$

Plugging into the previous equation, I get that:

$$\begin{aligned}\ln s_j^r - \ln s_0^r &= (x_j \beta_r - \alpha_r p_j + \xi_j) / \lambda + (\lambda - 1) (\ln s_t^r - \ln s_0^r) / \lambda \\ \lambda (\ln s_j^r - \ln s_0^r) + (1 - \lambda) (\ln s_t^r - \ln s_0^r) &= x_j \beta_r - \alpha_r p_j + \xi_j \\ \ln s_j^r - \ln s_0^r + (1 - \lambda) \ln \left(\frac{s_t^r}{s_j^r} \right) &= x_j \beta_r - \alpha_r p_j + \xi_j \\ \ln s_j^r - \ln s_0^r &= x_j \beta_r - \alpha_r p_j + (1 - \lambda) \ln \left(\frac{s_j^r}{s_t^r} \right) + \xi_j\end{aligned}$$

Notice that if $\lambda = 1$ (correlation within the nests = $1 - \lambda = 0$), the final equation reduces to the familiar multinomial logit market share expression. Notice also that s_j^r is not observed, only s_j is. Thus, it is not possible to identify different types, r , from

the regression approach; if I include more than one type on the RHS, the regression specification will exhibit perfect multicollinearity. For this reason, when using the 2SLS approach, I must restrict the model to only have one type of consumer.⁸ I proceed by estimating this regression equation using different instruments. In this way, I assess which instruments work best before extending the model to include random coefficients.

The reason I need instruments is because I allow for an arbitrary correlation between prices and the product-level unobservable attributes, ξ_j . In other words, $E(p_j \xi_j) \neq 0$. Similarly, the product's market share relative to the outside share is simultaneously determined with its within-nest share. Using the 2SLS approach above will make the estimates for β , α , and λ consistent, and allow me to test how well my instruments perform, by the following logic: the correlation between p_j and ξ_j will likely be positive, given that products with higher prices likely have more unobserved product attributes (first class ticket, last-minute purchase, etc.) that are desirable. For this reason, I expect products with higher prices to also have a larger error term, ξ_j , and a simple OLS regression would erroneously attribute higher utility when there are higher prices to be *caused* by the higher prices, making the fare coefficient, α , less negative. On the other hand, good instruments should not attribute the effect of demand-unobservables to the fare variable, making α more negative. I test for this, along with instrument relevance, in order to determine a set of instruments to use for the GMM estimation of the full structural model.

⁸In the next section, I will outline the GMM approach where I can back out both types.

Structural, GMM

Notice that from the observed market shares, s , I can back out the vector of product-unobservable characteristics in each market, t , denoted by ξ_t , as a function of the product characteristics, prices, the observed market shares and parameters:

$$\xi_t = s^{-1}(x_t, p_t, q_t, \theta_d) \quad (1.2.9)$$

As in Berry and Jia [2010], the multiple-type nested logit model requires a slight modification to the contraction mapping method used in BLP. In particular, the “step” between each iteration of the ξ_{jt} is as follows:

$$\xi_{jt}^M = \xi_{jt}^{M-1} + \lambda[\ln s_{jt} - \ln s_{jt}(x_t, p_t, \xi_t, \theta_d)] \quad (1.2.10)$$

Let z_t be a vector of instruments. Then the moment conditions used in estimation are based on restrictions of the form:

$$\begin{aligned} \mathbb{E}[\xi(x_t, p_t, q_t, \theta_d)|z_t] &= 0 \\ \Rightarrow \mathbb{E}[h(z_t)\xi(x_t, p_t, q_t, \theta_d)] &= 0 \end{aligned} \quad (1.2.11)$$

Then I can estimate θ_d using GMM.

I use the same instruments chosen in the regression analysis, and minimize:

$$\hat{\theta}_d = \arg \min_{\theta_d \in \Theta} \frac{1}{T} \sum_{t=1}^T h(z_t)\xi(x_t, p_t, q_t, \theta_d)$$

The instruments should include exogenous variables to help predict the endogenous characteristics (prices). Moreover, they have to identify parameters that govern substitution patterns across products in a market, such as the type specific parameters β_r , α_r , the nested logit coefficient λ , and the share of nonbusiness travelers γ_r .⁹ Exogenous variation in choice sets across markets will help to identify these substitution patterns,¹⁰ and I discuss instruments and identification in more detail after outlining the data I will use.

1.2.2 Supply

I assume that prices are set according to a static Nash equilibrium with multiproduct firms. Each firm, f , chooses the prices of its products, j in the set F , so as to maximize its profit subject to its capacity constraints. Let p be the vector of prices of all products, not just those produced by firm f , and let q^f be the vector of demands for products produced by firm f :

$$\begin{aligned} \max_{p_j} \pi^f &= \sum_{j \in F} (p_j - mc_j) q_j(p) \\ \text{s.t.} \quad & Bq^f(p) \leq c \end{aligned}$$

Where B is an $L \times F$ sparse matrix of 1's and 0's that capture the set of all products that use a particular link (F is the set of all products produced by firm f across all markets, t , and L is the set of all links operated by firm f ¹¹), and c is a vector of capacity constraints of length L . Recall that links are differentiated by their *operating* carrier, so

⁹These are all components of θ_d in the equations above.

¹⁰See Berry and Haile [2009] for a more formal argument, in a nonparametric context.

¹¹Recall that a particular product can use multiple links. Similarly, a link can be used by multiple products.

there could be more than one link between two cities. Letting b_{lj} be the lj -th element of the matrix B :

$$b_{lj} = \begin{cases} 1 & \text{if product } j \text{ uses link } l \\ 0 & \text{otherwise} \end{cases}$$

While there *are* fixed costs of operation in the short-run, an airline cannot change them so they will not affect the airline's optimal choice of prices. This chapter abstracts from the choice of network structure and capacity; in the appendix, I provide some ideas on the extension where airlines can endogenously choose their capacity.

Using μ_l to denote the Lagrange multiplier on a particular link, l , the first order condition of the above problem reads as follows:

$$q_j(p) + \sum_{r \in F} (p_r - mc_r) \frac{\partial q_r(p)}{\partial p_j} + \sum_{l \in L} \mu_l \sum_{r \in F} b_{lr} \frac{\partial q_r(p)}{\partial p_j} = 0$$

This equation captures the effects of changing the price of a product by one unit. Consider an increase of a dollar: first, it increases profit by the amount of passengers on that route, $q_j(p)$. Second, it lowers demand on that route, decreasing profit by the markup, $p_j - mc_j$, times the quantity change, $\frac{\partial q_j(p)}{\partial p_j}$. This relaxes the capacity constraints of all products that use this route. Third, this price increase may increase demand on all other products operated by this airline. This leads to two things: a positive change in profit on all other products (where this profit change is equal to the sum of all markups, $p_r - mc_r$, times the quantity change, $\frac{\partial q_r(p)}{\partial p_j}$), and a possible adverse effect on all products (except for the one whose price was increased), if those products have capacity-constrained

links ($\mu_l > 0$), caused by the tightening of constraints.¹² To simplify notation, let:

$$\Delta_{jr} = \begin{cases} -\frac{\partial q_r}{\partial p_j} & \text{if } r \text{ and } j \text{ offered by the same (ticketing) carrier} \\ 0 & \text{otherwise} \end{cases}$$

Then, I can re-write the above in matrix form and include the complementary slackness conditions to get the optimality conditions below. Notice that they differ from the standard supply-side model in Berry and Jia [2010] via the addition of the term with the Lagrange multiplier, and the complementary slackness condition. In the event that the constraints do not bind (i.e., $\mu_l = 0 \forall l$), we are back to the benchmark model in Berry and Jia [2010].

$$q - \Delta(p - mc) - \Delta B^T \mu = 0 \quad (1.2.12)$$

$$\mu^T [c - Bq] = 0 \quad (1.2.13)$$

To derive an analytic form for Δ , let Δ_1 and Δ_2 be the matrices of derivatives of market share with respect to mean utility, δ , for consumers of type 1 and type 2, respectively¹³. Then,

$$\Delta = \gamma \alpha_1 \Delta_1 + (1 - \gamma) \alpha_2 \Delta_2$$

Next, notice that $s_j = \frac{D^{\lambda-1}}{1+D^\lambda} e^{\delta_j/\lambda}$ and $\frac{\partial D}{\partial \delta_k} = \frac{e^{\delta_k/\lambda}}{\lambda}$

¹²There is a fourth effect that I do not explicitly write out here, for brevity, but that I take into account in simulation: marginal costs will be endogenous (they will depend on demand via an economy of scale variable), so changes in quantity will also change marginal costs.

¹³As before, if the products j and k are not offered by the same ticketing carrier, the jk -th element of the matrix is zero

$$\frac{\partial s_j}{\partial \delta_k} = \begin{cases} \frac{e^{\delta_k/\lambda}}{\lambda} \left(\frac{D^{\lambda-2}(-D^\lambda + \lambda - 1)}{(D^\lambda + 1)^2} \right) e^{\delta_j/\lambda} & : k \neq j \\ \frac{e^{\delta_k/\lambda}}{\lambda} \left(\left(\frac{D^{\lambda-2}(-D^\lambda + \lambda - 1)}{(D^\lambda + 1)^2} \right) e^{\delta_j/\lambda} + \frac{D^{\lambda-1}}{1+D^\lambda} \right) & : k = j \end{cases}$$

I can re-write the first-order condition as:

$$\ln(mc) = p - \Delta^{-1}(q + \Delta B^T \mu)$$

Since p , q , B as well as the matrix of partial derivatives, Δ , are from the data, if I know the multipliers, μ , I can solve for the implied marginal costs, mc . In order to perform counterfactuals, I need to project the implied marginal costs onto the observables (the *determinants* of marginal costs). For this reason, I specify a functional form for marginal cost. As in Berry and Jia [2010], let the marginal cost on route j be:

$$mc_j = \rho' w_j + \omega_j$$

The observable determinants of marginal cost, w_j include the following variables: the distance of the route, the distance of the route squared, the number of connections in the route, a dummy for whether the origin or destination is a hub for the operating carrier, a dummy for whether the product originates, ends, or passes through a slot-controlled airport, dummies for the different carriers, and the number of links in the route that are capacity constrained. The first three terms shifters of w_j , as well as the constant, are interacted with a dummy variable that distinguishes between short-haul and long-haul routes, since different aircraft are used for short-medium haul routes and long haul routes. The threshold for a long-haul route is a one-way distance greater than 1500 miles.

In other words, load factors and capacity enter in two places: 1) as a determinant of marginal costs, and 2) in the constraint. The first is indistinguishable from a specific form of economies of scale that arise once scale has reached near-capacity levels; in effect, my estimation procedure accounts for nonconstant marginal costs by allowing for a marginal cost “discount” once a certain load factor threshold is reached. The second is more interesting, and is used in estimating counterfactuals.¹⁴ With just economies of scale, an airline is incentivized to set (lower) prices such that load factors are high. With constraints, the airline is incentivized to set (higher) prices such that all links are operated within capacity. The interplay of these two effects is what distinguishes this paper from previous literature, and makes the answer nonobvious.

That said, in estimation, the capacity-constraint variable is endogenous since it depends on demand. For this reason, I assume that the cost-unobservables, ω_j , are mean-independent of some vector of observed instruments. I discuss instruments after describing the data.

1.3 Data

1.3.1 DB1B

The Airline Origin and Destination Survey (“DB1B”) is a 10% sample of airline tickets from reporting carriers collected by the Office of Airline Information of the Bureau of Transportation Statistics (“BTS”), for itineraries with *both* endpoints in the US. This data is aggregated **quarterly**.

¹⁴The reason I cannot use the constraints in estimation is due to the fact that they never bind in the data. In particular, since I only have aggregated monthly data, the total number of passengers is always less than the total number of seats. I discuss this further after describing the data.

The data is sorted into ticket, market, and coupon data. The ticket data provides itinerary-level information (fare, origin and destination, passengers). This ticket data is divided into markets, which are defined by a trip break, where a trip break is a point in the itinerary at which a passenger is assumed to have stopped for a reason other than changing planes. Note that this market definition *is not* the same as the market definition in this paper. Rather, my market definition is more closely related to the DB1B’s definition of itinerary. For example, an itinerary (ticket-level) BOS-LAS-BOS would have two markets: BOS-LAS and LAS-BOS. In this paper, a market is defined as a round-trip, e.g. BOS-LAS-BOS.

Finally, the market data is sorted into coupon-level data. A coupon is a piece of paper or series of papers indicating the itinerary of the passenger. For example, the market from BOS-LAS might include a layover in ORD. Thus, there would be two coupons for this market. As will be elaborated in the next section, a single coupon may involve a *stopover*, where the plane lands at an intermediate airport to deplane/enplane passengers but the flight number does not change.

1.3.2 T-100

The Air Carrier Statistics database, also known as the T-100 data bank (“T-100”), contains domestic and international airline market and segment data. To match the DB1B, I focus on the domestic airline market. This data is aggregated **monthly**.

In the T-100, a segment is defined as a nonstop flight between two points, while a market is defined as a “direct” flight between two points, where the flight number does not change. To clarify, a direct flight can contain a stopover, where the plane lands to

pick up and drop off passengers but has “continuing service” on to another airport. In fact, the *plane* can be changed at the stopover, but if the flight number remains the same, it still counts as a single market.

For this reason, the market data contains information on “on-flight market passengers emplaned,” while the segment data has the more informative “available seats” and “non-stop segment passengers transported,” along with characteristics of the plane. Also for this reason, matching the T-100 segment data to the coupon-level link data will lead to a nonnegligible portion of links from the coupon-level data (close to 25%) not being identified in the T-100 segment data. The reason for this is that these “links”, which had been implicitly assumed in the literature to be nonstop, were actually *direct*, and included a stopover. In this way, I can use the mismatch between the T-100 and coupon-level link data to generate a stopover dummy. The fraction of products with stopovers across airlines is consistent with expectation: Southwest has stopovers on 18% of its products, while Delta and United have half that, at 9%. American only has stopovers on 4% of its products.

1.3.3 Evidence for Load Factors

To provide support for the motivating statements on load factors outlined in the introduction, consider the following histograms of AA’s load factors for a similar period (the third quarter) before its merger with US (2012Q3) and after its merger with US (2015Q3). Define a “link” as any direct flight:

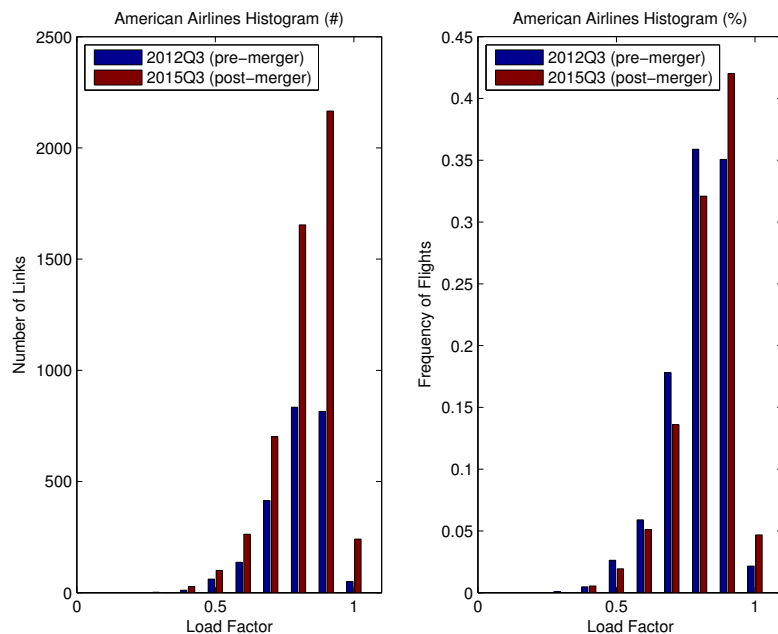


Figure 1.1: Histogram of American Airlines' Load Factors on Links

Notice that the *number* of links with a load factor between 90-100% increased by an order of magnitude from the pre-merger to the post-merger world (this can be seen by the last two bars on the left figure), while the *proportion* of links with a load factor between 90-100% more than doubled from the pre-merger to the post-merger world (this can be seen by the last two bars on the right figure). The difference is caused by the increase in number of links flown by AA after its merger with US. Either AA acquired more high-load links than low-load links from US, or the load factor of its links increased in the three years between 2012 and 2015, whether this was through creating/restructuring links or through an increase in demand — or a combination of any of the above.

This phenomenon is not limited to AA. Tables 1.1 and 1.2 show the number and fraction of links that are capacity constrained, for the 5 major airlines plus JetBlue, for the pre-merger period (2012Q3) and post-merger period (2015Q3). Note that all major

Table 1.1: Number of Links Operated $> 90\%$ Full

	2012 (Q3)	2015 (Q3)	% Change
Southwest	717	1937	170
Delta	1329	2307	74
United	1786	1956	10
American	349	1166	234
US Airways	681	-	-
JetBlue	117	198	69

Table 1.2: Percentage of Links Operated $> 90\%$ Full

	2012 (Q3)	2015 (Q3)	% Change
Southwest	13	26	109
Delta	24	35	46
United	26	30	14
American	15	23	51
US Airways	26	-	-
JetBlue	9	12	25

carriers increased the fraction of their routes that operated 90% full, or more,¹⁵ with Southwest more than doubling its fraction from 1/8 to 1/4.

1.3.4 Sample Selection

The sample selection procedures described below closely follow Berry and Jia [2010]. I estimate the model separately using data from the third quarters of 2012 and 2015. I report counterfactuals using estimates from 2012 data. As a result, I report summary statistics for 2012 data below, and delegate the summary statistics for 2015 data to the appendix. The sample selection procedure outlined below is almost identical for both samples.¹⁶

In the DB1B data, an itinerary is defined by a set of coupons that correspond to the individual flights in the journey. For example, an itinerary could be a round-trip between New York and San Francisco. If the flight connected in Chicago both ways, there would be a total of 4 coupons for this itinerary. As was mentioned earlier, this does not mean the plane flies *nonstop* between all of these cities; indeed, a direct flight from Chicago to San Francisco would count as an individual coupon even if it had a stopover in Salt Lake City, provided the flight number did not change. A market, on the other hand, is a directional pair of an origin and destination airport. The itinerary described above would have 2 markets — one from New York to San Francisco and the other from San Francisco to New York.

I drop the following itineraries: those that contain ground transportation (8.7%),

¹⁵The relevant measure here is more than 90% full, on average, over a month, since data is aggregated monthly.

¹⁶The only difference is that the top ticketing carriers that are kept in 2015 are different than those that are kept in 2012.

those that have more than 4 coupons (4.8%), those that have a self-loop (0.9%) (i.e., a non-ground transport connection within the same city, like a flight from LaGuardia to JFK), those that were not entirely *ticketed* by the same carrier (6.1%), those with abnormally low fares (5.6%) ($< \$25$, corresponding to, perhaps, the use of frequent-flier miles) or abnormally high fares (0.01%) ($> \$5000$), those with more than one stop in either direction (7.3%), those with more than two markets (1.1%), and those that are not round trips (32.6%). In total, this drops 46.9% of itineraries.

To define the size of each market, I use the geometric mean of the MSA population of the end-point cities. I focus on airports located in medium to large metropolitan areas with at least 700,000 people. I group together geographically close airports like Newark, LaGuardia and JFK. There were 4552 such markets in the third quarter of 2012. These markets roughly overlapped with the top 4000 most traveled markets, which is the scope of focus in many empirical studies. The reasons for excluding other markets are not only for computational reasons; there are also drastic differences in the nature of these markets, and they are often subsidized or operated at a loss. I keep these 4552 markets, leading to 30% of itineraries being dropped.

I only keep itineraries operated by the top 9 ticketing carriers, by passengers transported (dropping 10.5% of itineraries). These are, in order of passengers transported: Southwest, Delta, United, AA, US, JetBlue, AirTran Airways, Alaska Airlines, and Spirit Airlines. Note that I do not drop any itineraries by *operating* carrier — indeed, these 9 ticketing carriers will lead to a total of 23 operating carriers across products.

Finally, conditioning on observed characteristics, many itineraries have very similar fares and are not likely to be viewed by consumers as distinctive products. For this

reason, as in Berry and Jia [2010], I use progressive fare bins to aggregate records within each product.¹⁷ As a reminder, conditioning products on fare is important due to many product-unobservables that are correlated with fares (e.g. ticket class, advance purchase, date of travel). In summary, I define a product as a unique combination of origin airport, connecting airport(s), destination airport, operating carriers, ticketing carrier, and binned fare.

After all of these restrictions, my sample contains 594 thousand products and 4552 markets. This is over twice as many products as in Berry and Jia [2010], due to the fact that my supply-side analysis requires products be differentiated by their *operating* carriers on all legs.¹⁸ Overall, the entire sample selection procedure detailed above drops 65.4% of itineraries.

I present some summary statistics of my sample. First, I present summary statistics of the demand-side observables that are aggregated at the *itinerary*-level. This is to say, the averages are not passenger-weighted, but “itinerary-weighted”. Notice that the average fare ends up being close to \$500 this way. I define “Tourist Cities” to include New York City, Los Angeles, DC, and San Francisco. I define “Vacation/Resort Locations” to include Las Vegas, Atlantic City, and any airports in the Virgin Islands, Puerto Rico, Florida, and Hawaii. I define “Number of Direct Flights” as the number of cities the ticketing carrier serves directly from the origin.¹⁹ I define “Congested Airports” as those that are slot-controlled; this includes the airports LaGuardia, John F. Kennedy, Ronald Rea-

¹⁷In particular, I use the following set of bins: \$20 for all tickets under \$700, \$50 for tickets between \$700 and \$1000 and \$100 for tickets above \$1000.

¹⁸As a reminder, this is because two flights ticketed by the same carrier could have different load factors, depending on which airline is operating the flight. The T-100 dataset tells me the flow and capacity of direct flights by operating carrier.

¹⁹This variable is meant to capture a demand-side network affect akin to the positive utility of frequency in Wei [2014].

gan Washington National, and Chicago O’Hare. Since I group nearby airports together, this indicator effectively includes (all airports in) New York City, DC, and Chicago.²⁰ These variables are consistent with Berry and Jia [2010]. Notice that about 14% of links in my sample involve stopovers.

Table 1.3: Summary Statistics for the Observables (2012)

	Mean	SD	Min	Max
Connection	0.73	0.44	0	1
Fare (\$100s)	4.88	3.24	0.4	99
Tourist Cities	0.17	0.38	0	1
Vacation/Resort Locations	0.14	0.35	0	1
Number of Direct Flights	27.6	18.6	0	65
Distance (1000s miles)	2.62	1.48	0.13	11.8
Congested Ends	0.24	0.4	0	2
Congested Connections	0.25	0.58	0	2
Hub Origin	0.3	0.46	0	1
Number of Links w/ Stopovers	0.28	0.82	0	8

Below, I present some additional summary statistics, including the passenger-weighted fare. Notice how much lower the weighted fare is than the non-weighted one, at \$399. This means that, in our sample, the correlation between fare and number of passengers is negative. This suggests that larger groups of people are either better at finding lower fares (or more likely to get a group discount), or they tend to choose products with lower fares. The average number of products per market is 131 (again, this is more than twice what Berry and Jia [2010] had, due to the way I define products).

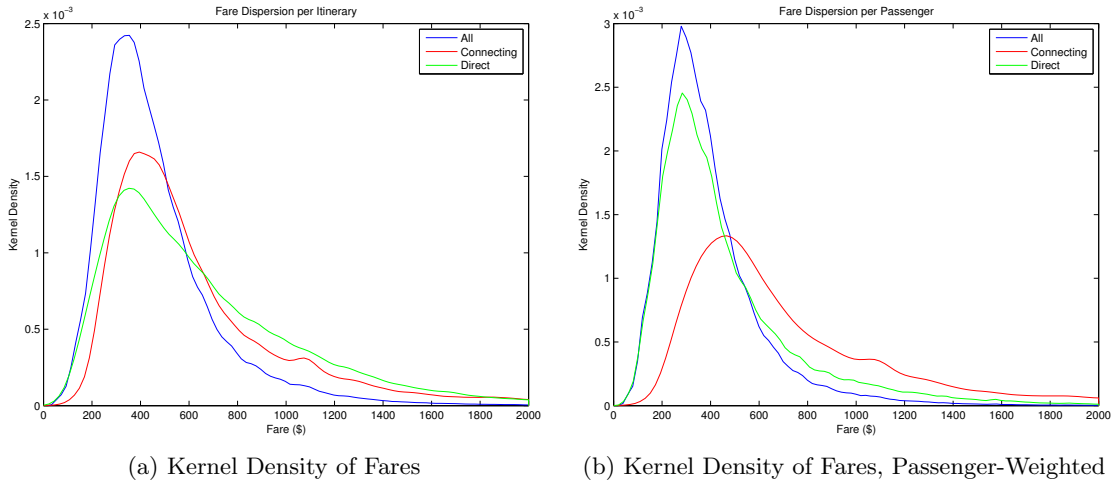
²⁰There is overlap with “Tourist Cities” defined above; only Chicago is uniquely in “Congested Airports” and only San Francisco and Los Angeles are uniquely in “Tourist Cities”.

Table 1.4: Other Summary Statistics (2012)

	Mean	SD	Min	Max
Products Per Market	131	165	1	1930
Unconditional Product Share (1e-5)	1.34	5.33	0.06	493
Conditional Product Share	0.01	0.02	0	1
Weighted Fare (\$100s)	3.99	2.58	0.5	100

To detail the differences in weighted vs. non-weighted fares further, I plot the distribution of these fares, and differentiate them by connecting or direct flights. Notice that the mean fare of connecting flights is higher than direct flights, and the distribution is more right-skewed. This is not surprising; while direct flights in the same market should be more expensive than their connecting counterparts, connecting flights should in general (unconditionally) be more expensive than direct flights. There are two reasons for this: 1) connecting flights may cover more distant markets, and/or 2) connecting flights may service more remote locations, where airlines' costs are higher.

Figure 1.2: Distribution of Fares, Non-Weighted vs. Weighted (2012)



Finally, I report summary statistics for the 9 ticketing carriers included in my sample. Notice that Southwest offers *fewer* products than Delta, but transports nearly 50% more passengers than Delta. Recalling that products are defined by route as well as by price bin and operating carrier, these summary statistics might suggest that Delta: 1) flies smaller planes, 2) has greater variance in fares, and/or 3) employs more operating carriers than Southwest. Knowing that Southwest is a low-cost carrier and largely operates its own flights, the third option seems to be particularly plausible. Indeed, this is the case. In our sample, Southwest operates *all* of the itineraries that it tickets. On the other hand, Delta uses 11 different operating carriers. As a result, if I remove product conditioning on operating carriers, Southwest has the same number of products (132 thousand) while Delta has much less (124 thousand).

Table 1.5: Summary Statistics for the Ticketing Carriers (2012)

	# Passengers	% Passengers	# Products	% Products
Southwest	740593	27.9	132090	22.3
Delta	536892	20.2	171343	28.9
United Airlines	392541	14.8	119454	20.2
American Airlines	314971	11.9	55987	9.5
US Airways	257097	9.7	82584	14
JetBlue	177153	6.7	8105	1.4
AirTran Airways	120028	4.6	15359	2.6
Alaska Airlines	69304	2.7	6280	1.1
Spirit Airlines	54359	2.1	2869	0.5

1.3.5 Adjustments

Cost of Fuel

While the cost of fuel is a significant driver of airline profits, it is impractical to include it as an explanatory variable in the marginal cost regression. The reason for this comes down to a data limitation: fare data is only available at the quarterly level. In other words, to get variation in fuel prices, I need to include multiple quarters of data in estimation.²¹ But this would introduce confounding effects arising from seasonality, unless I include seasonal dummies — these would not be separately identifiable from fuel prices. To get variation in both fuel prices and seasons, I would need to include many years of data. However, the product space would not remain constant, which is an important assumption of this paper. For this reason, existing literature has not included a fuel price explanatory variable (see, e.g., Berry and Jia [2010]).

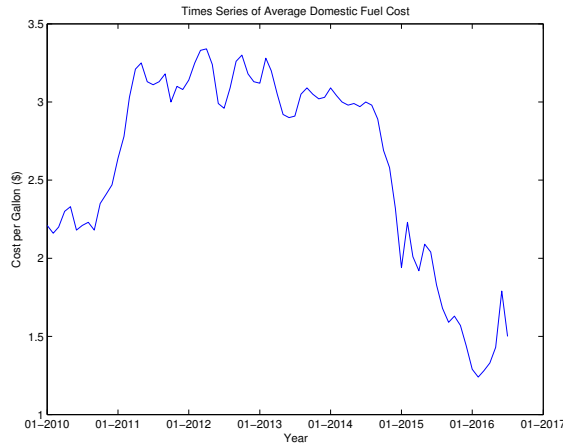


Figure 1.3: Average Domestic Fuel Cost per Gallon

That said, there is significant variation in fuel prices that should not be neglected,

²¹Recall that I only use a single quarter of data in estimation: the pre-merger period of 2012Q3.

both across carriers and across time. For example, the average spot price for jet fuel more than halved from the third quarter of 2012 (\$3.20 per gallon) to the third quarter of 2015 (\$1.55 per gallon) — these are the time periods of interest in this paper. Moreover, due to long-term contracts and hedging of spot prices, there was significant variation between the actual cost of fuel paid by different carriers. For example, in the third quarter of 2015, AA paid \$1.44 per gallon while Delta paid \$1.91 per gallon, a difference of 33%.²² Finally, due to differences in aircraft fleet composition and age, there is even more heterogeneity. For example, according to a 2014 report (Li et al. [2015]), airline efficiency on the A320 in 2014 varied from 8.1 passenger miles/lb fuel (Delta) to 11.3 passenger miles/lb fuel (Spirit), due to the average fleet age (2 years for Spirit and 19 years for Delta).

To get around the issue of estimating a cost of fuel parameter while still taking into account the drastic variation in fuel prices, both over time and across carriers, I perform the following accounting manipulation on the data. Using airlines’ 10-Q filings,²³ I calculate the carrier-quarter-specific fuel cost per revenue passenger mile (“FCPM”). This is given by the table below.

Table 1.6: Fuel Cost per Revenue Passenger Mile (cents/mile)

	2012Q3	2015Q3
American	5.77	2.91
Delta	5.96	3.34
JetBlue	5.31	3.09
Southwest	5.61	3
United	6.09	3.39

²²All data on fuel prices was obtained from the Bureau of Transport Statistics at <http://www.transtats.bts.gov/fuel.asp>.

²³Form 10-Q is a quarterly report mandated by the United States Securities and Exchange Commission (“SEC”) for all publicly traded corporations. 10-Q filings were obtained from the SEC’s website (www.sec.gov).

I then multiply the numbers in this table by the distance of each product and *subtract* this route-specific number from the backed-out marginal costs of every route. In other words, I eliminate from the backed-out marginal costs the portion of the “distance” explanatory variable that is due to estimated fuel costs. Then, when I forecast costs in a new period or in a counterfactual world, I *add* a new route-specific number derived by multiplying the numbers in this table by the distance of each product. This approach not only gets around the issues of including a cost of fuel variable, but it accurately captures heterogeneity in airlines’ cost of fuel and fuel efficiency. In the results section, I show that this approach leads to predictions for fuel costs that are consistent with airlines’ 10-Q filings. I also show that the distribution of fuel cost as a fraction of total variable cost is well-behaved across products.

In the merger counterfactual, I need to make an assumption on what FCPM would have been for AA and US, had they not merged (for the other carriers, I assume that FCPM is the same with or without the merger). I assume that AA has the same FCPM whether or not it merged with US, and I assume that US inherits the FCPM of the carriers that had the closest FCPM to it in the pre-merger period.²⁴ I find that my results are robust to different specifications, such as the case where AA’s FCPM is different if it had not merged (for example, if I assign AA the FCPM of carriers whose FCPM’s were closest to AA’s in the pre-merger period, by the same approach used to obtain FCPM for US in the relevant but-for world).

²⁴In particular, I solve for x_j such that the US FCPM in 2012 (which is 5.28) times x_j is equal to carrier j ’s FCPM in 2012. Then, I average over the top 5 carriers’ FCPM in 2015 times x_j .

1.3.6 Airfare Taxes

The fare reported in the DB1B data is the fare *paid by consumers*. In other words, it includes all taxes and fees. It is the appropriate fare to use in the demand-side of estimation, but not in the supply-side of estimation. Airlines do not actually receive this fare, but receive this fare minus all taxes and fees. That said, airlines may also receive ancillary fees relating to baggage and fuel surcharges, but I abstract from these on both the demand-side and the supply-side (partly because I am unable to identify whether a given fare has ancillary fees included, or not, e.g. premium economy vs. economy).

Taxes on *domestic* airfare fall into the following categories (using 2012 rates):

1. Domestic Passenger Ticket Tax: 7.5% of ticket price.
2. Domestic Flight Segment Tax: \$3.80 per segment.
3. Passenger Ticket Tax for Rural Airports: 7.5% of ticket price.
4. Flights between continental US and Alaska or Hawaii: \$8.40 tax per one-way trip.
5. Frequent Flyer Tax: 7.5% of value of miles.
6. Passenger Facilities Charges: up to \$4.50 (max of 4, for PFC-approved airports).
7. September 11th Security Fee: \$2.50 per enplanement (max of 2 per one-way trip).

In 2015, some of these taxes changed. In particular, the Domestic Flight Segment Tax rose to \$4.00; the Alaska/Hawaii Tax rose to \$8.90 per one-way trip; the September 11th Security Fee rose to \$5.60 (now, calculated as a total per one-way trip).²⁵

²⁵Data on airfare taxes was obtained from the Federal Aviation Administration's ("FAA"'s)

Below, I discuss the details regarding the implementation of these taxes. A rural airport is defined as one with fewer than 100 thousand enplanements during the second preceding calendar year, and either 1) is not located within 75 miles of another airport with more than 100 thousand enplanements, 2) is receiving essential air service subsidies, or 3) is not connected by paved roads to another airport. Since my analysis focuses only on MSA's with a population of over 700 thousand, I omit all "rural airports" from the analysis and can ignore this tax. The MSA of Anchorage is below 700 thousand, while the MSA of Honolulu is not. For this reason, all flights to and from Honolulu (and only those flights) have the additional \$8.40 tax added. I abstract from the Frequent Flyer Tax, since I do not have data on miles earned, and the miles earned can vary drastically between different types of Economy and Business fares (which I cannot differentiate from the data). A "Flight Segment" is defined in the same way a "Coupon" is in this paper: one takeoff and one landing. The Passenger Facilities Charge ("PFC") is only levied on PFC-approved airports, and the rate varies between \$3.00 and \$4.50.²⁶ To simplify analysis, I assume every airport levies a charge of \$3.50.

Putting all of this together results in the following total fare, F , as a function of the base fare, B , for a *round-trip itinerary*. Let nS be the number of segments:

$$F = 1.075 * B + 3.80 * nS + 6 * \max(nS, 4) + \mathbf{1}(\text{Honolulu}) * 16.80$$

website (https://www.faa.gov/about/office_org/headquarters_offices/apl/aatf/media/14.1.17excisetaxstructurecalendar2014.pdf) and from Delta's information on taxes and carrier-imposed fees (http://www.delta.com/content/www/en_US/traveling-with-us/planning-a-trip/booking-information/fare-classes-and-tickets/taxes-and-fees.html).

²⁶A list of PFC-approved airports was obtained from the FAA's monthly reports (https://www.faa.gov/airports/pfc/monthly_reports/media/airports.pdf).

In other words, given the DB1B fare date, F , I can use the following transformation as an approximation for the base fare airlines received, B :

$$B = 0.93 * (F - (3.80 * nS + 6 * \max(nS, 4) + \mathbb{1}(\text{Honolulu}) * 16.80))$$

The approach is as follows: I use F in demand-side estimation but B and F in supply-side estimation: e.g. $\Pi = (B - mc)q(F)$. I do the same when predicting counterfactuals, using the taxes in 2015 given above. To see this, notice that the first-order condition for profit maximization now reads as:

$$\begin{aligned} \max_p \pi_j &= \sum_j (f(p_j) - mc_j) q_j(p) \\ f(p_j) &= 0.93 * p_j - (3.80 * nS + 6 * \max(nS, 4) + \mathbb{1}(\text{Honolulu}) * 16.80) \end{aligned}$$

where p_j is the fare paid by consumers, and $f(p_j)$ is the fare received by airlines (the base fare). Then, the first-order condition of the above problem reads as follows:

$$f'(p_j) q_j(p) + \sum_{r \in J_j} (f(p_r) - mc_r) \frac{\partial q_r(p)}{\partial p_j} = 0$$

Rewriting the above in matrix form, I get:

$$0.93q(p) - \Delta(f(p) - mc) = 0 \tag{1.3.1}$$

Note that Δ is a function of p , not $f(p)$.

1.3.7 Instruments

In this paper, fares and capacity constraints are endogenous because of a classic simultaneous equations problem. In particular, prices and quantities are simultaneously determined in market equilibrium, causing fares to be correlated with the error terms in the demand function, and causing capacity constraints (via quantity) to be correlated with the error terms in the cost function. Fares are also endogenous due to an arbitrary correlation with the demand-side product unobservables, ξ_j .

Because of this endogeneity ($\mathbb{E}(p_j \xi_j) \neq 0$ and $\mathbb{E}(CC_j \omega_j) \neq 0$, where CC_j is a dummy for whether a particular product has capacity-constrained links), I need instruments in order to identify the fare and capacity-constraint coefficients. I discuss instruments for the fare coefficient first. Typically, demand studies will exploit the rival product attributes and the competitiveness of the market environment to instrument fares. But there is a problem with doing this here. Consider, for example, using the number of products as an instrument. This seems reasonable since, *ceteris paribus*, products with closer substitutes have lower prices. But this instrument may itself be endogenous because of the way products were defined; in particular, since a product is a group of tickets whose fares fall in a fixed bin, a market with wider price dispersion will have a larger number of products.

For this reason, I use route-level characteristics instead. Instruments along these lines are the number of rival carriers, the number of rival routes, the percentage of rival routes that offer direct flights, and the average distance of rival routes. A second group of (demand-side) instruments include cost-shifters that do not affect demand. Instruments along these lines are the number of cities to which a carrier flies nonstop flights from

the destination airport, and a dummy for connecting at a hub airport. A third group of instruments are the exogenous variables.

Next, I discuss instruments for the capacity-constraint variable. In addition to the instruments discussed above, following Wei [2014], I use two additional instruments for flow. The first instrument relies on the variation of “centrality” across links. In particular, it is the number of products that use the set of link l in a product. The second instrument relies on the variation in population across the endpoints of products. In particular, it is the geometric mean of the market size of the end cities of each product.

These instruments are relevant because a product is likely to have more capacity-constrained links if its links are used by many different products, or if they connect cities with large populations. They are valid because of the standard assumption that ξ is unknown to airlines before entry;²⁷ in other words, ξ is orthogonal to the network structure.

I also need instruments in order to identify the nested logit parameter, λ , and the share of nonbusiness travelers, γ . Gandhi and Houde [2015] show that there exists an ideal set of instruments in mixed logit demand systems that provide the fundamental source of variation in the data that identifies substitution patterns. They call these “Differentiation IV’s”; these instruments are aimed at capturing the distance in product space from similar products. One of my instruments captures the spirit of these Differentiation IV’s directly: “Rival Routes of Same Type”. This instrument measures the number of products that are connecting, if the product under consideration is a connecting flight, or the percentage of products that are direct, if the product under consideration contains a direct flight.

²⁷For a more elaborate discussion, see Aguirregabiria and Ho [2012].

Other instruments, while not exact matches, also aim to capture the distance in product space between similar products. I report summary statistics for some of the instruments used below:

Table 1.7: Summary Statistics for Selected Instruments

	Mean	SD	Min	Max
Rival Carriers	4.57	1.49	0	8
Rival Routes	17.4	13.1	0	85
Rival Routes of Same Type	12.4	11.9	0	79
% Rival Routes Direct	0.14	0.14	0	1
Distance of Rival Routes	2.88	1.53	0	12.5
Hub Destination	0.27	0.44	0	1
Own Routes	5.41	5.02	1	49
Link Centrality	705	550	0	3825

I perform instrument diagnostic tests following Nevo [2001]. First, I test for instrument relevance. To do this, I compute the F-statistic of a joint test of whether all instruments are significantly different from zero in the first-stage least-squares regression.²⁸ This statistic should be larger than 10, and it is in both cases. Second, I test for instrument validity. To do this, I perform the Sargan test on instrument validity (I can do this because the model is over-identified). I regress the residuals of the IV regression on the instruments and compare nR^2 from this regression to its distribution under the null of instrument exogeneity. The p-values for this test are 0 for both the price instruments and the capacity-constraint instruments, meaning that I cannot reject the null of instrument exogeneity. I conclude that my instruments are relevant and valid.

²⁸This is a partial F-test; it tests whether only the coefficients on the instruments are statistically different from zero, not whether all the coefficients are statistically different from zero.

1.3.8 Identification

I briefly outline identification of the nested logit parameter λ and the type-specific parameters, since the identification of the other parameters in the utility function is standard. Consider the identification of λ first. When $\lambda = 0$, the aggregate share of the routes remains fixed as the number of products varies across markets. In other words, the total demand in a market is inelastic. When $\lambda = 1$, the nested logit demand reduces to a simple logit. In other words, the total demand in a market becomes more elastic. As discussed earlier, I do not use the number of products as an instruments due to concerns about endogeneity.²⁹ That said, I use route-level characteristics, and the number of rival routes and carriers serve as a good proxy to the number of products in a market. Finally, I identify the type-specific parameters by exploiting substitution patterns among similar products when the mix of products varies across markets. For a more elaborate discussion of the above, see Berry and Jia [2010], Wei [2014], or the existing literature on random coefficient logit estimation.

1.3.9 Empirical Capacity Constraints

No link is capacity constrained, in the strict-sense, in the T-100 data.³⁰ The reason for this is that data is aggregated monthly, and planes always have a few empty seats over the course of a month. That said, it seems likely that a plane that had 99% of its seats full on average, over the course of a month, had at least a few individual flights that were completely full. Moreover, some flights may have had empty seats that were “off-limits”

²⁹Recall that products are conditioned on fare bin, and fare is an endogenous variable.

³⁰Throughout this paper, capacity and flows are aggregated at the monthly level. In other words, like the T-100 data, the predictions of flows measure the aggregate over a month, not an average per flight.

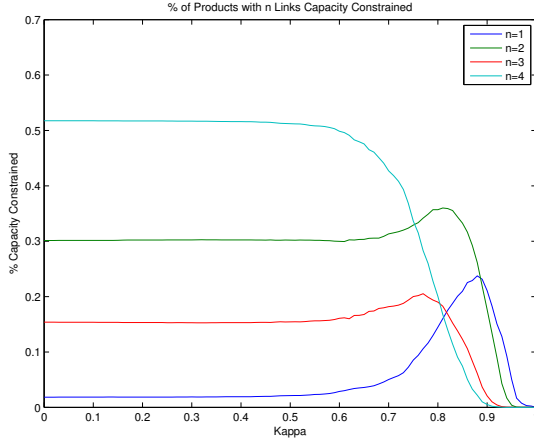
to passengers due to weight-balance considerations.

For this reason, it seems reasonable to assume that a link, j , is **effectively capacity constrained** ($\mu_j > 0$) if the monthly load factor on that link exceeds a threshold, $\kappa \leq 1$:

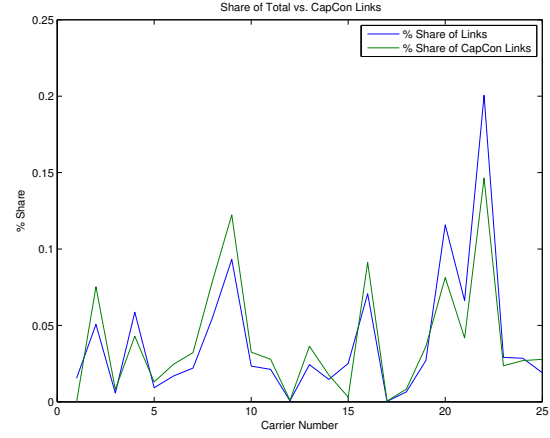
$$\frac{\sum_{s:s \in j} q_s}{c_j} \geq \kappa$$

There is a trade-off in selecting κ . On one hand, we want κ to be high enough to be meaningful, in the sense that it properly separates flights that are capacity constrained from flights that are not; on the other hand, we want κ to be low enough to include enough flights for there to be adequate variation to estimate the effect of capacity constraints accurately. In essence, we are turning a continuous variable (load factor) into a discrete binary variable (constrained or unconstrained).

To illustrate how the choice of κ affects the capacity-constraint variable, I plot two graphs below. The plot on the left shows the number of capacity-constrained links per product in my sample (recall that a product is a round-trip between an origin and destination, and I only keep trips with at most one stop), as a function of κ . The plot on the right takes $\kappa = 0.9$, and plots the share of capacity-constrained links vs. the share of total links per carrier. Notice that carriers, for the most part, have their “fair share” of capacity-constrained routes at $\kappa = 0.9$. Recall that our sample is from the third quarter of 2012.

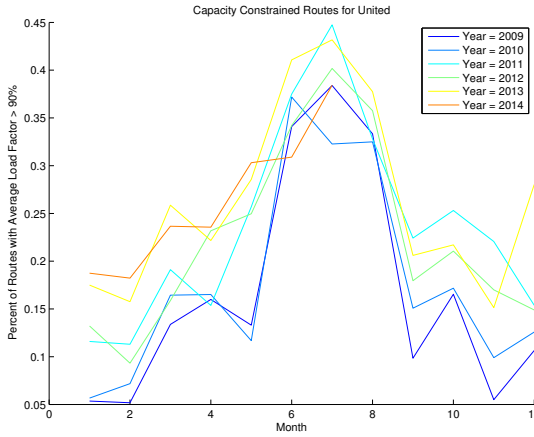


(a) % Capacity-Constrained Links vs. κ

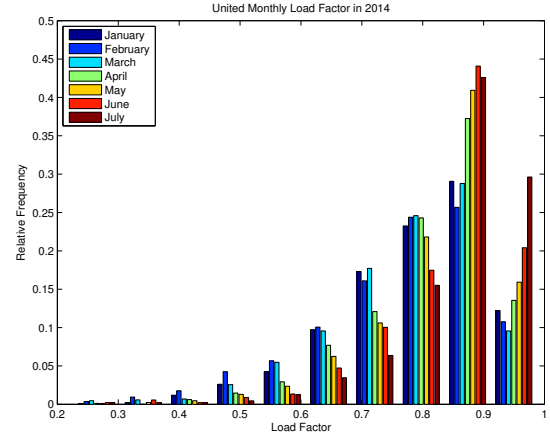


(b) % Capacity-Constrained Links at $\kappa = 0.9$

Below, I also plot the load factor and capacity-constrained routes (for $\kappa = 0.9$) for United over multiple years. Notice that there is seasonality in load factors: they are consistently higher in the summer and lower in the spring and fall. Throughout this paper, I will take $\kappa = 0.9$.³¹ In the results section, I will show how the estimates of the capacity-constraint coefficient vary with κ .



(a) Percentage of Capacity-Constrained Routes



(b) Histogram of Load Factors

I proceed as follows: I estimate the demand-side first, and then use the demand-side

³¹I find that my results are robust to taking κ to be 0.8, 0.85 and 0.94, and re-running the entirety of the analysis.

estimates to estimate the supply-side. I do not solve the full model in this paper; rather, I use the capacity information to generate a new variable, “capacity constrained” and include it in the supply-side marginal cost regression. Since this variable is endogenous, I instrument for it using the demand-side observables and instruments. I discuss extensions to the approach outlined above and present some ideas on how to solve the complete model, in the appendix.

1.4 Results

1.4.1 Demand

The BLP estimator minimizes an objective function in which instruments interact with both ξ_j and ω_j , and the demand-side and supply-side parameters are estimated jointly. In my application, this requires inverting the matrix Δ for many trial parameter values. Because of the very large product space in this paper,³² it is time-consuming to invert the matrix Δ . For this reason, I estimate the demand and supply side sequentially.

There are closed-form solutions for the price elasticity of demand in both the standard and nested logit models:

$$\frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = \begin{cases} \alpha p_j (1 - s_j) & \text{standard logit} \\ \frac{\alpha p_j}{\lambda} (1 - (1 - \lambda)s_{j|g} - \lambda s_j) & \text{nested logit} \end{cases}$$

Elasticities for all individuals (business, tourist) can be obtained by weighting individual elasticities by their relative shares: γ (for business travelers) and $1 - \gamma$ (for tourists).

³²Recall that this is driven by us having to distinguish products amongst operating carriers on each leg of the trip, in order to implement capacity-constraint analysis.

Elasticities for all products within a market can be obtained by weighting individual elasticities by their predicted market share (divided by sum of predicted market shares of all products within that market): $s_j / \sum_j s_j$. Elasticities for all products (across markets) can be obtained by weighting individual elasticities by their predicted market demand (divided by sum of predicted market demands of all products): $M_j * s_j / \sum_j M_j * s_j$. The weighting by market size is important here since some products will be in much larger markets than others (making the corresponding change in market shares different). I report elasticities for tourist and business travelers separately, as well as together, below.³³ I also calculate an “aggregate” price elasticity (as in Berry and Jia [2010]), where I find the percent change in total demand (sum of all products’ demands) when all prices increase by 1%. To make the coefficients in demand estimation comparable to each other, I compute a Willingness-To-Pay (“WTP”) measure by dividing each coefficient by the price coefficient. All carrier fixed effects are relative to US.

I estimate the model using 2012 data and 2015 data, separately. I present the demand side first. My estimates closely match Berry and Jia [2010], with all of the expected signs and magnitudes. As expected, the fare coefficient becomes more negative once I instrument for it.³⁴ I find that the price elasticity of demand of business travelers is 2.78, compared to 6.77 of their nonbusiness counterparts. The average price elasticity of demand (aggregated across types) is 5.15. The aggregate price elasticity of demand is 1.59. This number is not only consistent with Berry and Jia [2010],³⁵ but with a survey

³³These are calculated from the results of GMM with 2 types.

³⁴With standard OLS, we are led to erroneously attribute the effect of demand-unobservables to the fare variable. If these are positively correlated (as we expect, due to products with high fares likely having many utility-enhancing unobserved effects), it causes attenuation bias.

³⁵These authors found the aggregate price elasticity to be 1.55 in 1999 and 1.67 in 2006.

conducted by Gillen et al. [2003]; these authors collected 85 demand elasticity estimates from cross-sectional studies and found that elasticities ranged from 0.181 to 2.01, with a median of 1.33. I find that business travelers desire direct flights much more than their nonbusiness counterparts; in particular, the average business traveler would pay a 74% fare premium for a direct flight, vs. 32% for a nonbusiness traveler.

American enjoyed a sizeable “brand-name premium” in 2012; *ceteris paribus*, individuals would pay more to fly AA (over US) than to fly to a vacation/resort location (16% vs. 7% of average fare for the average nonbusiness traveler, and 40% vs. 17% for the average business traveler). On the other hand, individuals would need a lower fare to choose Southwest over US, *ceteris paribus* (19% for the average business traveler and 7% for the average nonbusiness traveler). However, AA’s brand-name premium disappeared in 2015; *ceteris paribus*, individuals would pay more to fly Delta or United over AA. AA still maintained its premium over Southwest.

In order to assess whether this change in brand-name premium was driven by AA acquiring lower-quality US products, I re-ran the above with two AA dummies: one for its legacy products, and one for its acquired products. I find that, in 2015, AA’s legacy products had a higher perceived quality than its acquired products from US (WTP about 50% higher). However, the brand-name premium of its legacy products still fell short of Delta’s or United’s products. For complete results, see the appendix.

The share of business travelers was estimated at 40%. This is consistent with Berry and Jia [2010], and if this seems high, recall that we implicitly assume individuals can only purchase one product or not fly. In other words, there is no measure of *frequency* of flying, and it may be the case that though business travelers are small in number, they

fly relatively more frequently than their tourist counterparts. The nested logit parameter is found to be 0.48 in 2012 and 0.45 in 2015. Berry and Jia [2010] find the nested logit parameter to be decreasing from 0.77 in 1999 to 0.72 in 2006. My results indicate that this trend has continued (and sped up) to 2012; in other words, airline products within a market have become closer substitutes over time. This likely reflects the reduced differentiation among products offered by different airline carriers as they cut down on services and competed more on prices.

The contribution of this paper on the demand-side is through the “stopover” variable. This variable keeps track of the number of stopovers in a product, where a stopover is defined as a direct flight that makes a stop to enplane/deplane passengers without the flight number changing. Recall that I was able to define this variable by comparing which coupons in the DB1B (which include direct *and* nonstop flights) did not appear in the T-100 (which include nonstop flights only). I find that the average individual needs about one-fifth the fare discount to purchase a flight with a stopover vs. a layover (14% for a stopover vs. 74% for a layover for the average business traveler, and 6% for a stopover vs. 32% for a layover for the average nonbusiness traveler). Previous literature would have considered flights with stopovers as direct flights. As can be seen from the estimates, stopovers generate a nonnegligible effect on consumer choice.

My estimates are almost all statistically significant. Following convention, three stars means significant at 1%, two stars means significant at 5%, and one star means significant at 10%.

Table 1.8: Demand-Side Parameter Estimates (2012)

	OLS (1-type)	IV (1-type)	GMM (1-type)	GMM (2-types)
Constant (Tourist)	-8.274*** (0.005)	-6.011*** (0.008)	-6.035*** (0.002)	-5.908*** (0)
Connection (Tourist)	-0.633*** (0.002)	-1.077*** (0.002)	-1.086*** (0.001)	-1.076*** (0)
Fare (Tourist)	-0.029*** (0)	-0.663*** (0.002)	-0.667*** (0.052)	-1.052*** (0.002)
Constant (Business)	-	-	-	-7.875*** (0.001)
Connection (Business)	-	-	-	-1.002*** (0)
Fare (Business)	-	-	-	-0.26*** (0.035)
Stopovers	-0.193*** (0.002)	-0.193*** (0.002)	-0.195*** (0)	-0.194*** (0)
Tourist Cities	0.13*** (0.002)	0.206*** (0.002)	0.207*** (0)	0.208*** (0)
Vacation/Resort Locations	0.516*** (0.002)	0.225*** (0.002)	0.226*** (0)	0.226*** (0)
Direct Routes	0.004*** (0)	0.013*** (0)	0.013 (2.646)	0.014 (2.399)
Distance	0.016*** (0.002)	0.204*** (0.002)	0.201*** (0.016)	0.206*** (0.018)
Distance ²	-0.007*** (0)	0.02*** (0)	0.02 (0.239)	0.02 (0.247)
Congested Ends	0.007*** (0.002)	-0.018*** (0.002)	-0.018*** (0)	-0.018*** (0)
Congested Connections	-0.046*** (0.001)	-0.087*** (0.001)	-0.088*** (0)	-0.087*** (0)
Hub Origin	-0.026*** (0.003)	0.072*** (0.003)	0.072*** (0)	0.072*** (0)
Southwest	-0.209*** (0.005)	-0.08*** (0.004)	-0.081*** (0)	-0.08*** (0)
Delta	-0.448*** (0.004)	0.143*** (0.004)	0.144*** (0)	0.144*** (0)
United	-0.432*** (0.004)	0.216*** (0.005)	0.216*** (0)	0.217*** (0)
American	-0.289*** (0.005)	0.537*** (0.005)	0.536*** (0)	0.542*** (0)
US Airways	-0.357*** (0.004)	0.17*** (0.005)	0.171*** (0)	0.171*** (0)
JetBlue	0.083*** (0.008)	0.165*** (0.007)	0.167*** (0)	0.165*** (0)
Lambda	0.518*** (0.001)	0.466*** (0.001)	0.463*** (0.079)	0.476*** (0.075)
Gamma	-	-	-	0.597*** (0.003)
Mu	-	-	0.099 (0.201)	0.102 (0.193)
Test of Over Identification (p-value)	-	0	-	-
1st Stage R^2	-	0.31	-	-
1st Stage F-test	-	159	-	-

Table 1.9: Demand-Side Parameter Estimates (2015)

	OLS (1-type)	IV (1-type)	GMM (1-type)	GMM (2-types)
Constant (Tourist)	-8.16*** (0.006)	-5.76*** (0.008)	-5.787*** (0.002)	-5.685*** (0)
Connection (Tourist)	-0.744*** (0.002)	-1.154*** (0.002)	-1.163*** (0.001)	-1.155*** (0)
Fare (Tourist)	-0.033*** (0)	-0.71*** (0.002)	-0.715*** (0.05)	-1.101*** (0.001)
Constant (Business)	- -	- -	- -	-7.659*** (0.001)
Connection (Business)	- -	- -	- -	-1.079*** (0.001)
Fare (Business)	- -	- -	- -	-0.307*** (0.035)
Stopovers	-0.263*** (0.004)	-0.37*** (0.004)	-0.373*** (0)	-0.369*** (0)
Tourist Cities	0.136*** (0.002)	0.256*** (0.002)	0.256*** (0)	0.256*** (0)
Vacation/Resort Locations	0.437*** (0.002)	0.117*** (0.002)	0.118*** (0)	0.118*** (0)
Direct Routes	0.007*** (0)	0.015*** (0)	0.015 (2.566)	0.015 (2.327)
Distance	0.035*** (0.002)	0.228*** (0.002)	0.225*** (0.015)	0.231*** (0.016)
Distance ²	-0.007*** (0)	0.025*** (0)	0.025 (0.214)	0.025 (0.221)
Congested Ends	-0.084*** (0.002)	-0.154*** (0.002)	-0.155*** (0)	-0.154*** (0)
Congested Connections	-0.03*** (0.001)	-0.043*** (0.001)	-0.043*** (0)	-0.043*** (0)
Hub Origin	-0.115*** (0.003)	-0.015*** (0.003)	-0.015*** (0)	-0.015*** (0)
Southwest	-0.389*** (0.005)	0.117*** (0.005)	0.118*** (0)	0.117*** (0)
Delta	-0.518*** (0.005)	0.728*** (0.006)	0.729*** (0)	0.736*** (0)
United	-0.579*** (0.005)	0.542*** (0.006)	0.542*** (0)	0.547*** (0)
American	-0.547*** (0.005)	0.336*** (0.005)	0.337*** (0)	0.338*** (0)
JetBlue	-0.073*** (0.008)	0.473*** (0.008)	0.477*** (0)	0.475*** (0)
Lambda	0.465*** (0.001)	0.554*** (0.001)	0.444*** (0.077)	0.454*** (0.071)
Gamma	- -	- -	- -	0.596*** (0.003)
Mu	- -	- -	0.099 (0.184)	0.101 (0.177)
Test of Over Identification (p-value)	-	0	-	-
1st Stage R^2	-	0.32	-	-
1st Stage F-test	-	1378	-	-

Table 1.10: Price Elasticity of Demand

	Tourist	Business	Total	ElasAgg
2012:	-6.77	-2.78	-5.15	-1.59
2015:	-7.78	-3.62	-6.17	-1.82

Table 1.11: Willingness to Pay (% of Average Fare in 2012)

	Tourist	Business
Connection	-32	-74
Stopover	-6	-14
Tourist Cities	6	15
Vacation/Resort Locations	7	17
Direct Routes	0	1
Distance	6	15
Distance ²	1	1
Congested Ends	-1	-1
Congested Connections	-3	-6
Hub Origin	2	5
Southwest	-2	-6
Delta	4	11
United	6	16
American	16	40
US Airways	5	13
Jet Blue	5	12
Average Fare (\$)	323	523

Table 1.12: Willingness to Pay (% of Average Fare in 2015)

	Tourist	Business
Connection	−31	−64
Stopover	−10	−22
Tourist Cities	7	15
Vacation/Resort Locations	3	7
Direct Routes	0	1
Distance	6	14
Distance ²	1	1
Congested Ends	−4	−9
Congested Connections	−1	−3
Hub Origin	0	−1
Southwest	3	7
Delta	20	44
United	15	32
American	9	20
Jet Blue	13	28
Average Fare (\$)	337	550

Once I have estimated the demand-side, I can back out what marginal costs are for each product. I compare the fuel component of this cost to airlines' 10-Q filings below:³⁶

Table 1.13: Fuel Component of Costs (% in 2012)

	Model	10-Q (low)	10-Q (high)
Southwest	51	39	56
Delta	33	28	37
United	44	39	50
American	44	37	51
US Airways	47	43	55

I also plot the distribution of the fraction of fuel cost to total variable costs by carrier.

³⁶I report an upper and lower bound for total variable costs. The upper bound is derived by taking the reported operating costs of each carrier and subtracting Depreciation and Amortization, as well as Special Items, if there were any. The lower bound is derived by taking the upper bound and further subtracting Salaries, Wages, and Benefits, since they could be interpreted as fixed costs in the short-run.

The distribution is well-behaved across products, for each carrier.

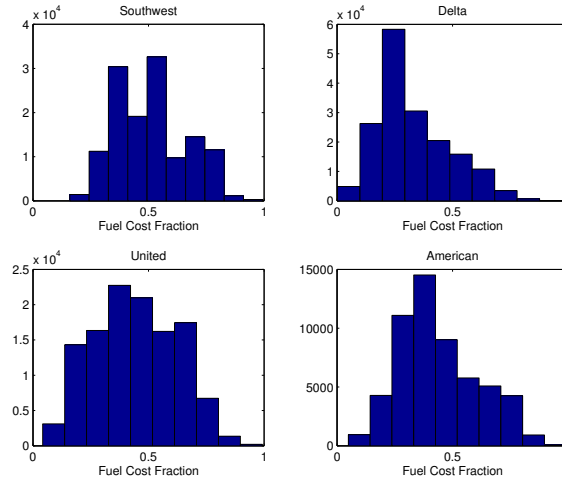


Figure 1.6: Distribution of Fuel Cost as a Fraction of Total Variable Cost

1.4.2 Supply

Next, I estimate the “reduced-form” version of the supply-side model, where capacity constraints enter as an additional explanatory variable for marginal cost. Note that the dependent variable, marginal cost, is in hundreds of dollars. I differentiate between short-haul (“SH”) and long-haul (“LH”) products the same way that Berry and Jia [2010] do: products with less than 1500 miles between origin and destination cities are SH, while products with more than 1500 miles between origin and destination cities are LH. 73.7% of products in my sample are LH.

I find that, for SH products, connections lead to lower marginal costs for the overall product (by \$21), while for LH products, connections lead to lower marginal costs for the overall product (by \$57). There are some intuitive reasons why making a stop would lower marginal costs: 1) it allows the plane to carry less fuel, and 2) it allows the airline to allocate better-suited planes to the demand of the route. There are also reasons why

a stop would increase marginal costs: it translates to an extra takeoff, landing, fuel/time spent taxiing and ascending/descending, as well as baggage handling and airport fees. It is reasonable that the cost savings discussed above affect SH routes more than LH routes. On the first point, carrying fuel for the entirety of a LH flight would translate to burning more fuel on route, due to extra weight, than for a comparable SH flight. On the second point, LH routes face constraints in the type/size of planes that can fly the direct distance; for example, it is not uncommon to see large, long-haul planes flying the relatively short ORD-JFK route, but there are no small jets flying the SEA-OSL route, even though the direct-route demand might warrant it. The length of this route requires a larger plane and, due to insufficient demand, a connection in a larger city or hub.

Including (and instrumenting for) capacity-constrained products, I find that, in 2012, the coefficient is significant and quite negative (-1.28).³⁷ This means that, *ceteris paribus*, capacity-constrained products have a lower marginal cost per passenger by \$128. I interpret this parameter as capturing the economies of scale that exist, from an airline's point of view, from having its flights near capacity. This is economically significant: the average marginal costs of products with a link that is operated close to capacity is less than half that of products without any such links (\$97 vs. \$202).

Consistent with Berry and Jia [2010], I find that JetBlue operates, *ceteris paribus*, at a much lower marginal cost than United — in particular, over \$180 lower. Marginal costs

³⁷The direction the coefficient moves after instrumenting is consistent with intuition. In particular, products with lower marginal costs are likely to be priced lower and hence demand a larger market share, making them more likely to be capacity constrained. This is a positive feedback effect on the capacity-constraint variable (higher flow leads to lower marginal cost which leads to even higher flow) that would cause OLS to underestimate the true effect of the capacity-constraint variable, in this case biasing it towards zero. Put another way, OLS erroneously attributes the variation in marginal cost to the final variation in the capacity-constraint variable, when it was likely a smaller variation in the latter that *caused* the variation in the former; the final (larger) variation in the latter is due to the feedback effect.

are increasing and concave in the distance of products: a decreasing amount of extra fuel and services are required for longer routes. As before, the distance variable is in thousands of miles.

My estimates are all statistically significant. As before, three stars means significant at 1%, two stars means significant at 5%, and one star means significant at 10%.

Table 1.14: Supply-Side Parameter Estimates (2012)

	OLS (no Cap Con)	OLS	IV
Short-Haul Constant	0.709*** (0.07)	0.714*** (0.07)	0.656*** (0.07)
Short-Haul Distance	1.344*** (0.14)	1.344*** (0.14)	1.685*** (0.14)
Short-Haul Distance ²	-0.333*** (0.07)	-0.332*** (0.07)	-0.248*** (0.07)
Short-Haul Connection	-0.213*** (0.01)	-0.213*** (0.01)	-0.208*** (0.01)
Long-Haul Constant	1.281*** (0.07)	1.259*** (0.07)	0.815*** (0.08)
Long-Haul Distance	0.203*** (0.01)	0.218*** (0.01)	0.91*** (0.02)
Long-Haul Distance ²	0.035*** (0)	0.034*** (0)	-0.014*** (0)
Long-Haul Connection	-0.575*** (0.01)	-0.573*** (0.01)	-0.565*** (0.01)
Hub (Origin, Destination, or Transfer)	0.268*** (0.02)	0.268*** (0.02)	0.131*** (0.02)
Congested Ends	0.169*** (0.01)	0.166*** (0.01)	0.023*** (0.01)
Capacity Constrained	- (0)	-0.02*** (0)	-1.28*** (0.03)
Southwest	0.161*** (0.02)	0.155*** (0.02)	-0.283*** (0.04)
Delta	0.583*** (0.02)	0.584*** (0.02)	0.667*** (0.04)
United	0.909*** (0.02)	0.908*** (0.02)	0.934*** (0.04)
American	0.96*** (0.03)	0.956*** (0.03)	0.535*** (0.04)
US Airways	0.612*** (0.02)	0.609*** (0.02)	0.442*** (0.04)
JetBlue	-0.085*** (0.03)	-0.098*** (0.03)	-0.884*** (0.04)
Test of Over Identification (p-value)	-	-	0
1st Stage R^2	-	-	0.3
1st Stage F-test	-	-	15

Table 1.15: Supply-Side Parameter Estimates (2015)

	OLS (no Cap Con)	OLS	IV
Short-Haul Constant	1.165*** (0.06)	1.184*** (0.06)	1.245*** (0.06)
Short-Haul Distance	0.279** (0.13)	0.268** (0.13)	0.454*** (0.13)
Short-Haul Distance ²	0.106* (0.06)	0.104 (0.06)	0.149** (0.06)
Short-Haul Connection	-0.18*** (0.01)	-0.184*** (0.01)	0.101*** (0.01)
Long-Haul Constant	0.4*** (0.06)	0.437*** (0.06)	0.364*** (0.06)
Long-Haul Distance	0.304*** (0.01)	0.269*** (0.01)	0.649*** (0.01)
Long-Haul Distance ²	0.031*** (0)	0.034*** (0)	0.005*** (0)
Long-Haul Connection	-0.535*** (0)	-0.545*** (0)	-0.306*** (0.01)
Hub (Origin, Destination, or Transfer)	0.361*** (0.02)	0.349*** (0.02)	0.327*** (0.02)
Congested Ends	0.059*** (0.01)	0.056*** (0.01)	0.042*** (0.01)
Capacity Constrained	- (0)	0.056*** (0)	-0.565*** (0.01)
Southwest	0.521*** (0.03)	0.51*** (0.03)	0.495*** (0.03)
Delta	1.278*** (0.03)	1.27*** (0.03)	1.623*** (0.03)
United	1.304*** (0.03)	1.287*** (0.03)	1.567*** (0.03)
American	0.927*** (0.03)	0.919*** (0.03)	1.262*** (0.03)
JetBlue	0.581*** (0.03)	0.606*** (0.03)	0.521*** (0.03)
Test of Over Identification (p-value)	-	-	0
1st Stage R^2	-	-	0.26
1st Stage F-test	-	-	615

In order to assess if there are any differences in costs between AA's legacy products

and acquired products in 2015, I ran the above specification with two AA dummies: one for its legacy products, and one for its acquired products. I find that, in 2015, AA’s legacy products have almost the same cost fixed-effect (“FE”) as its acquired products from US (\$125 vs. \$120, respectively). For complete results, see the appendix.

Because the value of κ is assumed, it is prudent to examine how the model’s results would vary as κ changes. Positive values of the capacity-constraint coefficient, for example, would cast doubt on the reliability of the model. To this end, I re-run the above regression for different choices of κ , and plot the resulting capacity-constraint coefficient. This is, in effect, a plot of the reduction in marginal costs due to a flight being capacity constrained, given different definitions of “capacity constrained”. I find that the reduction in marginal costs due to a flight being capacity constrained is U-shaped in the choice of κ , and always the same sign. This captures the spirit of the this paper: for low values of κ , economies of scale dominate, causing marginal costs to decrease as we increase the capacity-constraint threshold, while for high values of κ , capacity constraints dominate, causing marginal costs to increase as we increase the capacity-constraint threshold. In other words, as the makeup of capacity-constrained products moves towards those flights with very high load factors, we lose out on the gains from economies of scale due to the “costs” of capacity constraints binding. Of course, this increase in marginal costs should not be taken literally — what is really happening here is that observed *prices* are higher for capacity-constrained products once we make κ high. This lends credence to the importance of there being a capacity constraint in the airline’s pricing problem; when load factors are very high, airlines want to increase prices on those routes in order to prevent the constraint from binding.³⁸ I

³⁸As mentioned in the previous section, while I take $\kappa = 0.9$ for the remainder of the paper, results are robust to taking κ to be 0.8, 0.85, and 0.94. The figure above illustrates that these choices of κ cover most

illustrate this in the figure below.

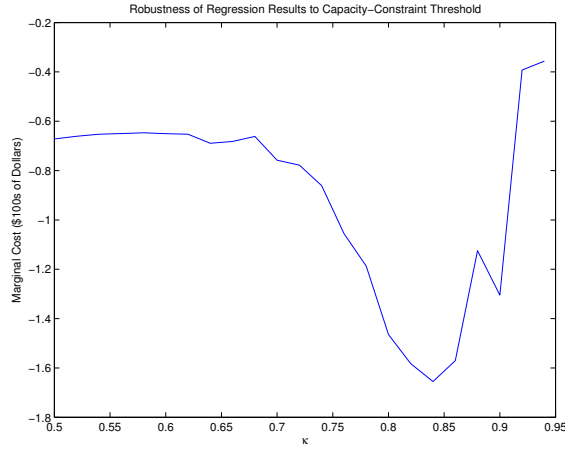


Figure 1.7: Reduction in Marginal Costs vs. κ

Finally, while some of the observations in the dependent variable in the cost regression may be negative, due to the fact that I am subtracting an average cost of fuel from products, almost no marginal costs are negative in prediction.³⁹ Almost all negative instances of the dependent variable get absorbed in the error term, as per the plot below.

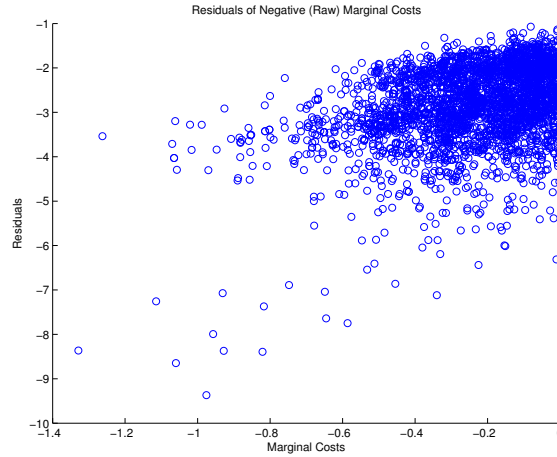


Figure 1.8: Negative Marginal Costs Fully Explained by Residuals

of the range of values the capacity-constraint coefficient can take.

³⁹In particular, only 0.05% are.

1.4.3 Prediction

In order to perform an analysis of model fit and counterfactuals, it is necessary to predict quantities, costs, and prices from the model. Doing so requires the assumption that preferences and the determinants of costs do not change after a merger. Given the results in the previous section, it is clear that preferences and the determinants of costs *did* change. As a result, I perform the model fit and counterfactual analysis using (third-quarter) 2012 and 2015 estimates, separately. I find that my predictions are relatively robust to either specification. Consequently, in the model fit and counterfactual sections that follow, I report results using 2012 estimates.⁴⁰

To predict costs, I use the estimated cost function. To predict prices, I minimize the optimal pricing equation (under Bertrand competition) under some norm:

$$\hat{p} = \underset{p}{\operatorname{argmin}} ||q(p) - \Delta(p)[p - mc(p)]|| \quad (1.4.1)$$

Since there are products in 2015 that did not exist in 2012, this approach requires predicting the unobserved components of utility, ξ , and costs, ω , for certain products. To do this, I take the set of ξ 's and ω 's obtained from estimating the model using 2012Q3 data, and place them in price bins⁴¹ and a capacity-constrained bin. Then, for each new product in 2015, I integrate over the conditional distribution of ξ and ω corresponding to that product's price bin, and whether it is capacity constrained. In this way, I can predict costs and demands for new products in 2015 without using 2015 estimates. This

⁴⁰Moreover, I prefer to use 2012 estimates over 2015 estimates because it allows me to estimate fixed-effects for US when it is not merged.

⁴¹The price bins used here are the same as those used when defining products.

is consistent with ex-ante merger analysis, where the resulting costs of new products are now known prior to the merger in question.

Recall that marginal costs are endogenous: prices will affect quantities which could change which links are capacity constrained. As discussed previously, since I need to distinguish products by their operating carrier on each link, the product space in this paper is very large (594 thousand in the third quarter of 2012). In other words, this problem amounts to solving a very large system of equations, where at each step of the minimization, I need to construct a large amount of entries in the (sparse) derivative matrix. That said, I can exploit the monotonicity of the FOC in prices to accelerate convergence, as follows:

1. Initialize a vector of prices, $p^N = p^0$, for all products
2. Calculate $\Delta(p^N), q(p^N)$
3. Calculate $mc(q(p^N))$ (depends on q through the capacity-constraint variable)
4. Set a vector of new prices using the first-order conditions: $p^{N+1} = mc - \Delta^{-1}q$
5. Increase prices on products that have links where flow exceeds capacity
6. Stop when $\|p^N - p^{N+1}\|$ is sufficiently small and all constraints are satisfied; go to Step 2 otherwise

The third step allows me to capture economies of scale, and the fifth step allows me to capture the effect of capacity constraints.⁴² Notice that these are distinct steps, and

⁴²The fifth step continues until no links are capacity constrained, in the sense that flow exceeds capacity. These links could still be “capacity constrained” in the sense that they capture the economies of scale coefficient.

in the counterfactual section I will show what happens as either is omitted. Note that the observables must be exogenously determined in any counterfactual, since product structure and entry/exit is not endogenously determined by the model. In particular, the structure of the matrix $\Delta(p)$ may differ from one counterfactual to the next due to changes in product ownership/space. For robustness, I try multiple product-space configurations.

While this approach attempts to disentangle the effects of economies of scale and capacity constraints, it should be made clear that doing so is never fully possible. In particular, since, in estimation, the capacity-constraint variable captures both economies of scale and capacity-constraint effects, even if I shut down the capacity constraints themselves (by allowing the flow on links to exceed capacity), the (negative, capacity-constraint) effects of operating close to capacity will still be present in the economies of scale variable that affects marginal costs.

1.4.4 Model Fit

The focus of this paper is on assessing the fit of models with and without capacity-constraint considerations, in the context of mergers. For this reason, it is natural to assess model fit in the following way: to what extent do the forecasts of the pre-merger estimates (cost function and preferences) match the realized equilibrium in the post-merger world? I define “equilibrium” by the set of prices, quantities, costs, and profits, and “match” the percentage change in these objects after the merger. I calculate the following models: the benchmark model without capacity constraints (“No CC”), the model with exogenous capacity constraints (“Exog. CC”),⁴³ and the model with endogenous capacity

⁴³By exogenous capacity constraints, I mean that the set of products that are capacity constrained in the post-merger period (2015Q3) are obtained directly from the data; they do not enter into the airline’s

constraints (“Endog. CC”). I present results for the 4 major airline carriers and JetBlue, in the table below:

Table 1.16: Estimated CPM (cents/mile)

	No CC	Exog. CC	Endog. CC	10-Q (low)	10-Q (high)
Southwest	9.1	7.2	6.9	6.9	12.3
Delta	8.6	9.9	10.9	10.4	14.3
United	9.6	12.2	12.4	9.3	13.8
American	8	8.8	9.2	10.6	15
JetBlue	9.5	7.5	8.1	7.8	11.3

Table 1.17: Estimated Variable Profit Margin (%)

	No CC	Exog. CC	Endog. CC	10-Q (low)	10-Q (high)
Southwest	37	59	61	23	121
Delta	37	45	45	14	56
United	36	36	36	14	68
American	38	40	40	14	61
JetBlue	35	50	52	24	80

For all model predictions, I use the pre-merger estimates (2012Q3) to predict the post-merger world (2015Q3).⁴⁴ In particular, to calculate costs per passenger mile (“CPM”), I multiply the predicted marginal costs (using 2012Q3 estimates) of each product⁴⁵ (in 2015Q3) by the number of passengers purchasing that product, and divide this result by the distance of the product multiplied by the number of passengers using that product. In pricing decision.

⁴⁴I assume that AA’s fixed effects are the same in the post-merger world as they are in the pre-merger world. I relax this assumption in the counterfactual section.

⁴⁵I do not make any assumptions on how capacity-constrained links have changed; I find which links were actually capacity constrained in the post-merger world when predicting marginal costs. In other words, I make use of all observables here. As outlined in the previous section, I predict demand and costs for new products by integrating over the conditional distribution of ξ and ω using 2012Q3 estimates.

other words, my CPM estimate is a weighted average of marginal cost per mile of products (since some products have many more passengers than others). To calculate profit margins, I multiply the markup (fare minus predicted marginal cost) of each product by the number of passengers purchasing that product, and divide this result by the predicted marginal cost of that product multiplied by the number of passengers purchasing that product. In other words, my profit margin estimate is a weighted average of markups as a percentage of cost. I use actual fares (when calculating profit margins) and actual miles flown.⁴⁶

I compare the cost estimates from my model to the income statements of the 4 major airlines' and JetBlue's 10-Q filings for the first quarter of 2015. It is important to note that while my estimation focused exclusively on the domestic market, the income statements in the 10-Q filings do not distinguish between domestic and international operations. Prior to the merger, AA stated in its 2013Q3 10-Q filing that about 60% of its passenger revenues were derived from domestic operations. Given that the majority of revenues were likely from the domestic market, the usefulness of these benchmarks should not be neglected. See Nevo [2001].

When calculating CPM and profit margins from the 10-Q filings, I report an upper and lower bound. For CPM, the upper bound is derived by taking the reported operating costs of each carrier and subtracting Depreciation and Amortization, as well as Special Items, if there were any. The lower bound is derived by taking the upper bound and further subtracting Salaries, Wages, and Benefits, since they could be interpreted as fixed

⁴⁶To clarify, the above is a partial equilibrium approximation; I do not simulate a full equilibrium with a corresponding set of prices here, though I do so when discussing fares below, as well as in the counterfactual section.

costs in the short-run. Then, I divide both by the total number of revenue passenger miles (taken directly from the 10-Q filings). For profit margins, I sum operating revenues from Mainline Passengers and Regional Passengers, and divide these by the upper and lower bounds of CPM (before they’ve been divided through by revenue passenger miles) to obtain upper and lower bounds.

Notice that the model with endogenous capacity constraints outperforms the other models. In particular, for the CPM estimates, the model with endogenous capacity constraints is within the 10-Q bounds for every carrier except AA, while the model without capacity constraints is outside the 10-Q bounds for Delta and AA. The model with exogenous capacity constraints is outside the 10-Q bounds for JetBlue, as well. For the profit margin estimates, all models are within the 10-Q bounds for every airline.

Table 1.18: Estimated Median Fares

	No CC	Exog. CC	Endog. CC	Actual Median
Southwest	396	341	352	432
Delta	518	472	481	489
United	549	509	504	487
JetBlue	462	459	437	450
American	531	509	483	472

Notice, from Table 1.18, that the models with capacity constraints outperform the model without capacity constraints for most airlines. In particular, for most airlines, the median fares predicted from the models with capacity constraints are closer to the actual median than those predicted from the model without capacity constraints. It is difficult to tell whether the model with endogenous or exogenous capacity constraints performs better; the model with exogenous capacity constraints is closer for JetBlue, while the

model with endogenous capacity constraints is closer for all other carriers.

There is a straightforward explanation for the superior model fit obtained by including capacity constraints. Notice that the capacity-constraint variable has a negative effect on costs in the pre-merger world. In other words, links operating closer to their capacity have a lower marginal cost per passenger. The CPM estimates in the model with capacity constraints are often lower than their counterparts in the model without capacity constraints. If the number or composition of capacity-constrained links dramatically changed after the merger, the influence of this variable could cause predicted costs to be different than in a model where we omitted capacity-constraint considerations. In particular, if the number of capacity-constrained links increased substantially, predicted costs could be lower in a model that included a capacity-constraint variable in the cost regressions. Indeed, as outlined in the introduction to this chapter, this is the case.

Having established that there is good model fit, and that the results can easily be explained with the motivating data, I turn to the analysis of counterfactuals.

1.4.5 Counterfactual: But-For AA-US Merger

The first counterfactual I simulate is a world where there was no AA-US merger, *ceteris paribus*. AMR Corporation, the parent company of AA, and US Airways Group, the parent company of US, merged on December 9, 2013. They continued separate operations, and it was not until April 8, 2015 that the Federal Aviation Administration granted a single operating certificate for both carriers.

There are a few things that the merger might do in practice in the short-run. First, the merger will change ownership structure. In this section, I use 2012Q3 (the “pre-merger

period”) estimates and construct hypothetical ownership structure in 2015Q3 (the “post-merger period”), had the merger not occurred, by matching the observable characteristics of products in the post-merger period to those in the pre-merger period, in a procedure outlined below.

Second, the merger may affect the cost and perceived quality of products that changed ownership. For example, when AA acquires US’ PHL-PHX route, they may continue to fly the route with US’ fleet, pilots, and personnel, and continue to use US’ terminal areas, at least in the short-run. Alternatively, AA may substitute its own fleet, pilots, and personnel on the route, and use its own terminal areas in airports where it had operational overlap with US. If the latter occurs, we may think that AA’s costs would be higher since, in 2012, AA had a higher cost FE than US. Evidence from estimating the model in 2015 using 2 dummies for AA’s products suggests that cost fixed-effects were very similar for AA’s legacy and acquired products in 2015.⁴⁷ For this reason, throughout the counterfactual section, I assume that the cost fixed-effects of products that changed ownership (from US to AA) attain the cost fixed-effects of the pre-merger AA.⁴⁸

Similarly, when AA acquires US’ routes, those routes may experience an increase or no change in perceived quality. If AA successfully rebrands these products as its own, they may be able to attain the higher quality FE of AA over US. If it is not able to (in the short-run), say, because the routes are still operated with US’ fleet, pilots, and personnel, or continue to use US’ terminal areas, these acquired products may retain their pre-merger

⁴⁷See the appendix for complete results; acquired products have a slightly lower cost FE.

⁴⁸I use pre-merger fixed-effects throughout all simulations of post-merger worlds. The reason I do this is because I rely on 2012 estimates throughout my counterfactual section. While I do estimate the model using 2015 data, I cannot mix and match the observed fixed-effects in 2015 with my other estimates in 2012. I prefer to use 2012 estimates over 2015 estimates because it allow me to estimate fixed-effects for US when it is not merged.

quality. Evidence from estimating the model in 2015 using 2 dummies for AA’s products suggests that demand fixed-effects (quality) were higher for AA’s legacy products than for its acquired products, in 2015.⁴⁹ For this reason, throughout the counterfactual section, I report complete results for two specifications: one where the quality of products that changed ownership (from US to AA) retain their original pre-merger quality FE, and one where they attain the quality FE of the pre-merger AA.

Finally, the merger may affect the product-specific qualities and costs (ξ and ω). I assume that these effects are second-order, and give all products that existed in the pre-merger world their pre-merger ξ and ω . For new products in the post-merger world that did not exist in the pre-merger world, I integrate over the conditional distribution of ξ and ω corresponding to that product’s price bin, and whether it is capacity constrained, as discussed earlier.

Having discussed what the merger might do in practice in the short-run, I turn to a discussion of how I determine product ownership in the hypothetical but-for merger (“BFM”) worlds.

Simulating Product Ownership

My goal in this section is to change as little as possible in the post-merger product space, apart from the ownership of products. In particular, I want to allocate all of AA’s post-merger products to either (a fictional, independent) US or AA in 2015. Most merger analysis has proceeded in the opposite direction: mergers are simulated by taking the pre-merger product space and combining ownership of two carriers’ products into a single,

⁴⁹See the appendix for complete results; legacy products have an over 50% higher FE.

unified carrier. The problem I face here is more difficult: given the post-merger world, which of the unified carrier's products would have belonged a fictional AA or US but-for the merger?

There are two main reasons why one might be interested in analyzing mergers ex-post: to learn whether enforcement has been too permissive or too strict, or to evaluate and potentially improve economic models used in antitrust enforcement. Indeed, Hosken et al. [2016] show that ex-post merger evaluation has helped improve methodology and reinvigorate enforcement towards hospital mergers in the US. In an ex-post environment, performing merger counterfactuals using only pre-merger data throws away valuable information about the change in product structure during the merger period (in this case, 2012Q3-2015Q3). For example, two airlines went out of business during this time period, for reasons likely exogenous to the merger in question. Since this paper abstracts from entry/exit in the product space (the only endogenous variable is the price of existing products), using ex-ante data would erroneously keep these two airlines in operation in the post-merger period. The methods developed in this section provide a robust way to perform ex-post merger evaluation with models that have exogenous network structure. As a simplification, I assume all of AA's products in 2015 would have been operated by either US, AA or both in the but-for world. To motivate an iterative product-assignment algorithm, consider the tables below:

Table 1.19: Intersection of Products

	US(2012)	AA(2012)	US+AA(2012)	AA(2015)
US(2012)	81646	0	0	0
AA(2012)	0	55987	0	0
US+AA(2012)	0	0	137633	0
AA(2015)	3111	18220	21331	183761

Recall that, in this paper, products are characterized by their route (including the origin, destination, and stops), fare bin, and operating carrier on all legs. In the third quarter of 2012, US had close to 82 thousand products, while AA had close to 56 thousand products. In the third quarter of 2015, AA had 184 thousand products. Notice that it does not make much sense to match products at this level: most of the products offered by US were operated by US, leading to a very low intersection with products operated by AA in 2015 (only 21 thousand products). Even if we correct for this (labeling US-operated routes as AA-operated routes), the resulting table is not much better:

Table 1.20: Intersection of Products (Alternate)

	US(2012)	AA(2012)	US+AA(2012)	AA(2015)
US(2012)	82584	0	0	0
AA(2012)	550	55987	0	0
US+AA(2012)	0	0	138021	0
AA(2015)	17643	18220	35331	183761

There is a slight increase in the products that US had in 2012, because we are now distinguishing products operated by US in 2012 from products operated by other non-US Airways carriers that were also not present in 2015. Doing away with the outdated “US”

operating-carrier label, there are nearly twice as many product overlaps from 2012 to 2015 (21 vs. 35 thousand out of a total of 184 thousand products in 2015), but this still amounts to less than a fifth of products in 2015. The reason for this is partly driven by rebranding: PSA Airlines, previously a subsidiary of US, now flies under the American Eagle brand for AA, and operates as “Envoy” after 2014. Moreover, Chautauqua Airlines, a regional airline, and AirTran Airways, a low-cost airline, ceased operations or were acquired by parent companies and rebranded. Moreover, any differences or dispersions in fares will lead to a different or augmented product set. Indeed, it makes more sense to go a step higher, and look at the intersection of *routes*, instead of products:

Table 1.21: Intersection of Routes

	US(2012)	AA(2012)	US+AA(2012)	AA(2015)
US(2012)	11180	0	0	0
AA(2012)	336	8306	0	0
US+AA(2012)	0	0	19150	0
AA(2015)	7579	5685	12955	27341

There is still a 50% increase in routes from the pre-merger to the post-merger world (19 thousand operated by AA and US together in 2012, to 27 thousand operated by AA in 2015). To understand this phenomenon, consider the market from Knoxville, TN to Tulsa, OK. In 2012, US and AA offered two routings for this round-trip market; one stopped in Chicago, IL on the way to the destination and Dallas, TX on the way back, while the other stopped in Dallas in both directions. In 2015, AA offered four routings for this round-trip market: two the same as in 2012; one that stops in Dallas on the way to the destination and Charlotte, NC on the way back; another that stops in Charlotte

in both directions. Charlotte used to be a hub for US; the merger allowed AA to double the amount of routes offered on the Knoxville to Tulsa round-trip market via the addition of a single hub to its network. This effect is present in many of the routes in the post-merger world and should help to explain the doubling of routes from the pre-merger to the post-merger worlds.

Notice that there is very little route overlap between US and AA in 2012, while there is significant overlap between US in 2012 and AA in 2015, as well as AA in 2012 and AA in 2015. This tells us that the merged AA combined and kept the routes flown by AA and US in 2012.

Table 1.22: Intersection of Markets

	US(2012)	AA(2012)	US+AA(2012)	AA(2015)
US(2012)	3138	0	0	0
AA(2012)	2085	2977	0	0
US+AA(2012)	0	0	4030	0
AA(2015)	3091	2948	3957	4182

While there are many unassigned routes, there are only a little over 200 unassigned origin and destinations (“markets”). In particular, US and AA together operated 4030 markets in 2012, of which 3957 coincided with AA’s 4182 markets in 2015. In other words, we can account for many of the previous “missing” routes, by matching those routes according to which airline (US or AA) had control of the underlying market. Finally, we could look at origins only:

Table 1.23: Intersection of Origins

	US(2012)	AA(2012)	US+AA(2012)	AA(2015)
US(2012)	68	0	0	0
AA(2012)	64	65	0	0
US+AA(2012)	0	0	69	0
AA(2015)	68	65	69	69

Out of the 70 possible MSA’s used in our sample (recall that these are MSA’s with a population exceeding a population of 700,000), AA flew from 65 in 2012 and 69 in 2015. In particular, it added flights from 4 new cities (Allentown, Albany, Bakersfield, and Sarasota), from which US was already flying in 2012. It is reasonable to assume that all products originating in these 4 cities should be allocated to US in a world but-for the merger. On the other hand, US never operated flights from the McAllen-Edinburg-Mission MSA. It is reasonable to assume that all products operated from this MSA in 2015 by AA would have also been operated by AA in 2015 but-for the merger.

With that exposition in mind, consider the following iterative algorithm. The general idea is to start at the lowest level (product-level), and match AA’s 2015 products to either AA in 2012, US in 2012, or both, where possible. Out of the remaining products, go one level up (e.g. to the route-level, by removing the fare and operating carrier from products) and match the remaining unassigned products where possible to either AA in 2012, or US in 2012, or both. At some point (and this point can be adjusted for robustness), start giving AA “priority”, in the sense that if AA matches with the remaining products at some level, it gets those products, and US matches with what’s left — as opposed to giving it to both carriers if they both match (as this creates extra products and we want

to keep this to a minimum; we know that if we kept this up, both carriers would match on most of the origins). It also seems reasonable that most of the new routes opened in the world but-for the merger would have been opened by AA, as it was the surviving company. If anything, US may have closed some of its routes due to low profitability. Giving AA “priority” after some point in the iterations attempts to account for this.

Formally, call the set of products that AA offered in 2012 $A2_1$. $A2_1$ is characterized by the following characteristics: $A2_1 = \{O, D, S_T, S_B, F, C_T, C_B\}$, which stand for the following:

- O : Origin airport
- D : Destination airport
- S_T : Stop to the destination, if any
- S_B : Stop back, if any
- F : Fare bin
- C_T : Operating carriers on flights to the destination
- C_B : Operating carriers on flights back

Similarly, define $U2_1$ and $A5_1$ for the set of products offered by US in 2012, and AA in 2015, respectively. Next, define the subsets of $A2_1$, $U2_1$, and $A5_1$ as follows:

- $A2_2 = \{O, D, S_T, S_B, F\}$
- $A2_3 = \{O, D, S_T, S_B\}$
- $A2_4 = \{O, D\}$

- $A2_5 = \{O\}$

The subset $A2_2$ is the set of unique products offered by AA in 2012 if we do not keep track of operating carriers. $A2_3$ is the set of unique routes flown by AA in 2012. $A2_4$ is the set of unique markets serviced by AA in 2012. Finally, $A2_5$ is the set of unique origins AA flew from in 2012. Similarly, we define the subsets of $U2$ and $A5$. Additionally, let $A5_i^R$ denote the remaining (unassigned) products in $A5$ at any subset-level i .

Let j be the point at which we start giving “preferential treatment” to AA, in the sense that at this iteration and all consequent iterations, we first match AA’s products in the BFM world, and remove those from the remaining products, before matching to US’ products. The algorithm proceeds as follows:

Iterative Route-Assignment Algorithm

1. Initialize $i = 1$ and $A5_1^R = A5_1$
2. Find the intersection of products (or subsets of products): $A = A2_i \cap A5_i$ and $U = U2_i \cap A5_i$
3. If subsets of products, project A and U to $A5$ to get all products associated with A and U . Store.
4. Find the remaining (unassigned) products: $A5_i = A5_i \setminus (A \cup U)$
5. Project the remaining products onto the next subset: $A5_{i+1}^R$
6. Repeat steps 2-4 until $i \geq j$

7. Then, find the intersection of subsets: $A = A2_i \cap A5_i$
8. Find the intersection of *remaining* subsets: $U = U2_i \cap (A5_i \setminus A)$.
9. Repeat steps 3-5 and 7-8 until $i = 5$.

The results of the algorithm are displayed below, for $j = 3$.

Table 1.24: Visualization of Product Allocation (n=183761)

	AA	US	Doubled	Remaining
Iteration 1	0	0	0	0
Iteration 2	32956	42948	4428	112285
Iteration 3	27933	42022	2441	44771
Iteration 4	39507	4153	0	1111
Iteration 5	1018	93	0	0
Summary	101414	89216	6869	-

Robustness Check 1

While retaining all the existing products in 2015, the approach detailed above does away with some products operated in 2012. In particular, we might be inclined to think that, but-for the merger, US would have continued to operate the products that it operated in 2012. As a first robustness check, I add products corresponding to the 2012 *routes* of US to its 2015 BFM products derived via the algorithm above. The reason why I match on the route-level, as opposed to the product or market-level, is detailed below:

Table 1.25: Products Remaining from 2012

	US(2012)	US(2015BFM)	Overlap	Remaining	Remaining Products
Product	82584	89216	17643	64941	64941
Route	11180	9489	7577	3603	9275
Market	3138	2941	2906	232	788

The first row tells us that only 17 thousand of the nearly 88 thousand products assigned to a hypothetical US in 2015 overlapped exactly with US' products in 2012. This is not surprising, given that that we distinguish products by their operating carrier, and is very similar to the product overlap in Table 16, when we do not distinguish between products operated by AA or US. The reason for the small difference (17643 vs. 17111) is because there were some products that AA operated in 2015 that were exactly the same as what it operated in 2012, and almost the same as what US operated in 2012 (save the operating carrier being US). I give these products to AA and not both.

If we tried to take set differences at the product level, we would be left with over 65 thousand additional products that US operated in 2012 that were seemingly cut by AA in 2015. Of course, this is not the case. Apart from changes in operating carriers over time, there are changes in fares. Removing these two product characteristics, the second row tells us that were some 3603 *routes* cut by AA after the merger, which translated to 9275 fewer products. For this robustness check, I match on this level. In other words, I give the products corresponding to these routes back to a hypothetical US in 2015. This leads to an extra 9275 US products in estimation in the 2015 BFM world.

The third row of the table tells us that there are very few markets (232) that US flew

in 2012 that were cut by AA after the merger in 2015. Moreover, these markets only translated to a few products (788). This lends some credibility to only looking at 2015 products.

Robustness Check 2

As a second robustness check, I calculate counterfactuals in the “traditional” way — i.e., I take the 2012 products as the 2015 BFM products for all carriers, and simulate the effects of joining products (and deleting duplicated products) from a joint AA carrier. I define a “duplicated product” not necessarily at the product-level. In particular, I want to remove the fare and operating carrier components for products. The reason for this is as follows: 1) after the merger, AA will reoptimize prices; I do not want to artificially increase the product space if AA and US were flying identical products but-for the price in 2012, and 2) many operating carriers (including US) merged or went bankrupt by 2015; I do not want to artificially increase the product space if AA and US were flying identical products but-for the operating carrier in 2012.

Recall from the table above detailing the intersection of routes that there was very little route overlap between US and AA in 2012, while there was significant overlap between US in 2012 and AA in 2015, as well as AA in 2012 and AA in 2015. This tells us that the merged AA likely combined and kept the routes flown by AA and US in 2012. One approach, then, would be to create new products according to *route-level* differences.

Table 1.26: Robustness Check: Merging US and AA in 2012

	US(2012)	AA(2012)	Overlap	Difference	Added Products
Product	82584	55987	0	82584	82584
Route	11180	8306	336	10844	77407
Market	3138	2977	2085	1053	22257

This approach gives AA an additional 77 thousand products out of a possible 83 thousand products that US operated in 2012. This is similar to the change in number of products in the first two counterfactual simulations, and I follow it in the next section. The results are largely robust to the level of matching.

Results

Below, I simulate the model *abstracting from any capacity considerations*. What I mean by this is that I simulate the model where the capacity-constrained links do not endogenously change after the merger (i.e., they remain at their pre-merger level) and there are no capacity-constraint considerations (i.e., the firm faces an unconstrained maximization problem). The reason I simulate this model first is that it provides a useful benchmark for the next section and allows easier comparisons amongst the different robustness checks outlined in the previous section.⁵⁰ As discussed before, I use 2012 estimates throughout this section.

⁵⁰In the next section, when I introduce capacity considerations, I have to link capacity data from the T-100. The robustness checks outlined above use 2012 data in addition to 2015 data, which means that I need to use multiple years of T-100 data and find a way to project capacity from 2012 to 2015. This clouds the comparability of the robustness checks, since I do not model capacity change. For example, consider a route operated by Delta in 2012 that doubled in capacity by 2015. R2 would more frequently have this route capacity constrained, since it relies on 2012 observables, while R1 and R2 would more frequently have this route operated below capacity. While I do scale for aggregate changes in capacity, I cannot capture idiosyncratic changes. I leave the endogenous capacity problem for future work.

In the first column of each table, I present the predicted value of the variable in question in the post-merger period using the observed data (“Predicted”, i.e., with a merger). In the second column, I show the percentage change between the post-merger period, and a hypothetical BFM world. In other words, this column shows the percentage change in the variable of interest that can be attributed to the merger. In the third and fourth column, I show the percentage change between the post-merger and but-for world, where the but-for world has been simulated using robustness check 1 (% Change R1) or robustness check 2 (% Change R2), respectively.⁵¹

First, I present results where only the ownership of products changes. In other words, I present results where there is no “quality transfer” in the sense that AA’s acquired products from US attain the same demand FE as AA’s legacy products. Rather, I assign the AA pre-merger demand FE to AA’s legacy products, and the US pre-merger demand FE to AA’s acquired products from US. After presenting these results, I will present results that include quality transfer in addition to the change in ownership of products.

⁵¹Note that the values determined via robustness check 2 are not compared to the values in the “Predicted” column. Rather, to determine the percentage changes under this robustness check, I calculate different merged values (i.e. in the same vein as the “Predicted” column) and compare them to the de-merged values under robustness check 2. Refer to the robustness check section for more details.

No Quality Transfer

Table 1.27: Counterfactual CPM (cents/mile)

	Predicted	% Change	% Change R1	% Change R2
Southwest	9.1	0	0.3	0.1
Delta	8.6	-0.1	-0.2	-0.1
United	9.6	0	0.2	0
American	8	0	-0.7	0.2
US	10	8.9	12	8.4
JetBlue	9.5	-0.1	0.6	0.2
AA+US	8.7	1.4	1.5	4.1

I find that the merger increased CPM for the merged airline (“AA+US”) by 1.4-4.1%. This is driven by the products that AA inherited from US: those products become significantly more expensive once they get rebranded as AA products and, as a consequence, acquire the AA cost FE.⁵² While the costs for products operated by other carriers do not change, the relative weight of passengers using those products does, as market shares change endogenously in response to the merger. As a result, the overall CPM of all carriers changes slightly.

⁵²If I do not give the AA cost FE to products that AA acquired from US, and instead allow those products to retain their US cost FE, this phenomenon does not occur, and the combined AA+US airline’s costs are virtually unchanged because of the merger.

Table 1.28: Counterfactual Median Fares

	Predicted	% Change	% Change R1	% Change R2
Southwest	396	-0.4	-0.2	-0.3
Delta	518	-0.1	0	-0.1
United	549	-0.2	-0.1	-0.1
American	531	2	2.1	1.7
US	534	7.2	5.9	8.4
JetBlue	462	-0.4	-0.4	-0.4
AA+US	533	4.9	4.3	6.2

Table 1.29: Counterfactual Variable Profit Margin (%)

	Predicted	% Change	% Change R1	% Change R2
Southwest	37	-0.2	0.2	-0.4
Delta	37	-0.3	-0.7	-0.4
United	36	-0.2	0.1	0.4
American	38	3.1	3.2	3.2
US	44	5.5	7.2	2.2
JetBlue	35	-0.4	-2.7	0
AA+US	40	3.4	3.7	2

The merged airline increases fares on its products by an average of 4-6%, but passenger volume actually decreases for products that were acquired from US. There is a 3% increase in variable profit margins on the merged airline's legacy products, and a slightly higher increase in variable profit margins on the merged airline's acquired products from US (2-7%). However, driven partly by the volume drop on acquired products, the resulting variable profit margins for the merged airline are only 2-4% higher because of the merger.⁵³

⁵³The results in the right-most column are not a typo. While AA's variable profit margins increased by 3.2% on its legacy products and 2.2% on its acquired products, AA's total variable profit margins did indeed only increase by 2%. This is akin to Simpson's paradox: the denominators are not the same in both groups. This number does not report the change in profits because of the merger, but the change in profit *margins*. This phenomenon happens in some of the other variable profit margin tables that follow,

When I include quality transfer, in the next section, variable profit margins for the merged airline will increase.

Table 1.30: Counterfactual Summary of Merger: All Carriers

	Predicted	% Change	% Change R1	% Change R2
Median Fares	498	1.6	1.3	1.5
Passengers (000s)	1770	−1	−3.1	−1.4
RPM (cents/mile)	12.9	1.1	1.5	1.4
CPM (cents/mile)	8	0.4	0.8	1.1
Variable Profit Margin (%)	38	1.1	1.2	0.5
Products (000s)	648	−0.8	−2.2	−0.9
Consumer Surplus	0	−2.2	−4.5	1.2

The table above tells us that the merger may have hurt or harmed consumers: consumers surplus was between 1.2% higher to 4.5% *lower* because of the merger. At the same time, it increased average variable profit margins by 0.5-1.2%. Put another way, the AA-US merger likely hurt consumers but helped airlines — specifically the merging airline, whose variable profit margins increased by 2-4%.

as well as one of the median fare tables. The reason for the phenomenon in the latter is different from the reasons here; it is due to the nature of the median function.

Quality Transfer

Table 1.31: Counterfactual CPM (cents/mile), Quality Transfer

	% Change	% Change R1	% Change R2
Southwest	−0.2	0.1	−0.2
Delta	−0.4	−0.5	−0.2
United	−0.1	0.1	−0.1
American	0.1	−0.6	0.7
US	7.3	10.4	6.7
JetBlue	−0.3	0.5	−0.1
AA+US	3.5	3.5	3.5

When I allow for quality transfer, the merger increased CPM for the merged airline by 3.5%. This is again driven by the products that AA inherited from US: those products become significantly more expensive once they get rebranded as AA products and, as a consequence, acquire the AA cost FE. The reason these numbers differ from those where there is no quality transfer is because of a change in the relative weight of passengers that use products. For example, when there is quality transfer, US products have a smaller increase in costs because of the merger (7-10% vs. 8-12%). This implies that quality transfer causes the relative weight of passengers to shift to lower-cost US products.

Table 1.32: Counterfactual Median Fares, Quality Transfer

	% Change	% Change R1	% Change R2
Southwest	−1.1	−0.9	−1.1
Delta	−0.7	−0.5	−0.8
United	−0.4	−0.3	−0.5
American	2.6	2.7	2.6
US	9.7	8.4	10.8
JetBlue	−1.4	−1.3	−0.6
AA+US	6.6	6.1	7.6

Table 1.33: Counterfactual Variable Profit Margin (%), Quality Transfer

	% Change	% Change R1	% Change R2
Southwest	−1.1	−0.7	−1.5
Delta	−1.4	−1.8	−1.3
United	−0.6	−0.3	−0.2
American	3.7	3.8	4.1
US	6.6	8.3	5.4
JetBlue	−1.1	−3.4	−0.2
AA+US	5.7	6.1	5

The merger caused variable profit margins for the merged airline to increase by 5-6%. This number is more consistent with reality: AA’s third quarter 2015 report claims that: “The Company’s third quarter 2015 pretax margin excluding net special charges was [...] up 6.7 percentage points from the same period last year.” This result is driven partially by giving products acquired from US the same demand FE (quality transfer) as AA products.⁵⁴

To understand the mechanism, consider what happens when there is a quality transfer. In this case, products previously operated by US experience a large increase in passenger

⁵⁴Recall that US had a lower demand FE than AA.

volume after the merger (10-23%), driven by higher demand by rebranding these as AA products. This allows the merged airline to increase fares on products acquired by US by 8-11%, and enjoy higher variable profit margins on these products by 5-8%. By virtue of being a larger airline with more market power, AA is also able to increase fares on its legacy products (2.7%) and enjoy higher variable profit margins on its legacy products (3-4%). Overall, the merged airline enjoys 5-6% higher variable profit margins as a result of the merger.

Table 1.34: Counterfactual Summary of Merger: All Carriers, Quality Transfer

	% Change	% Change R1	% Change R2
Median Fares	1.9	1.6	1.4
Passengers (000s)	3.7	1.7	3.2
RPM (cents/mile)	2.6	2.9	2.4
CPM (cents/mile)	1.4	1.7	1.9
Variable Profit Margin (%)	1.9	2	0.7
Products (000s)	-0.8	-2.2	-0.9
Consumer Surplus	5.1	2.9	2.7

The table above tells us that the merger increased consumer surplus by 3-5%. At the same time, it increased average variable profit margins by 1-2%. Put another way, the AA-US merger helped both consumers and airlines — specifically the merging airline, whose variable profit margins increased by 5-6%. In other words, welfare gains due to the merger can be attributed to quality transfer driven by the positive rebranding of US products to AA products. In the next section, I show that the positive welfare effects of the merger can persist even absent this quality transfer, if we allow for capacity constraints.

Effect of Capacity Constraints

Next, I simulate the full model to incorporate and highlight the role of capacity constraints. Throughout the results below, I use 2012Q3 estimates and the observed product structure in 2015. As discussed earlier, I predict costs and demand for products in 2015 that were not in 2012 by integrating over the conditional distribution of ξ and ω . To simulate the but-for world, I rely on the benchmark iterative route-assignment algorithm outlined in the counterfactual section that split AA's 2015 products between a hypothetical, de-merged AA and US in 2015. In this section, I introduce capacity in two distinct ways: 1) I allow marginal costs to be endogenous, allowing airlines to strategically capture economies of scale if demand exceeds κ times capacity, and 2) I do not allow flow to exceed capacity, forcing airlines to satisfy their capacity constraints. As discussed in the prediction section, simulating the model in this case involves finding a fixed point where all constraints are satisfied.

In the first table, I show the CPM of all carriers in two BFM worlds. The first column is the complete model using 2012 data, where marginal costs are endogenously determined in simulation, and there are capacity constraints. The second column is the partial model, without capacity constraints. All models are compared to the actual 2015 world where there is no merger.⁵⁵ In essence, the second BF world attempts to capture a model with economies of scale: airlines reoptimize their routes such that many are capacity constrained and operated at a lower marginal cost. The full model (the first BF world) limits the extent to which airlines can do this: there is a trade-off with increasing

⁵⁵I have allowed capacity constraints to be different in the actual 2015 world than what was observed in order for results to be comparable.

flows on routes, since it may cause the capacity constraint to bind.

First, I present results where only the ownership of products changes. In other words, I present results where there is no “quality transfer” in the sense that AA’s acquired products from US attain the same demand FE as AA’s legacy products. Rather, I assign the AA pre-merger demand FE to AA’s legacy products, and the US pre-merger demand FE to AA’s acquired products from US. After presenting these results, I will present results that include quality transfer in addition to the change in ownership of products.

No Quality Transfer

Table 1.35: Counterfactual CPM (cents/mile), Full Model

	Predicted	% Change 2012	% Change 2012 (EoS)
Southwest	6.9	0.3	−0.1
Delta	10.9	6.9	4.9
United	12.4	0.8	0.8
American	9.2	−5.1	−6.4
US	5.9	1.2	2.2
JetBlue	8.1	−1.8	0
AA+US	7.9	−1.4	−1.6

Table 1.36: Counterfactual Capacity Constrained Links, Full Model

	Predicted	% Change 2012	% Change 2012 (EoS)
Southwest	400.8	−1	0.1
Delta	100	−17.3	−10.3
United	26.5	−2.3	0.1
American	51.8	20.8	20
US	83	4.9	−1.8
JetBlue	13.6	4.6	0
AA+US	134.8	11.4	7.2

In general, CPM is driven by changes in the number of capacity constrained links before/after the merger (the “EoS” effect). In particular, notice that Delta loses 10-17% of its capacity constrained links because of the merger — this translates into a cost increase of 5-7%. On the other hand, AA’s legacy products become 20-21% more capacity constrained because of the merger — this translates into a cost decrease of 5-6%. It is likely that the merged airline competes most directly with Delta; Southwest and United also have a decrease in number of capacity constrained links, but much less than Delta (1-2%).

Table 1.37: Counterfactual Median Fares, Full Model

	Predicted	% Change 2012	% Change 2012 (EoS)
Southwest	352	0.1	−0.5
Delta	481	1.2	0.8
United	504	0.3	0
American	483	0.4	−0.1
US	475	4.2	4.3
JetBlue	437	−4.9	0
AA+US	479	2.4	2.1

The difference in cost predictions between the two BFM worlds translates to different median fares, as well. In particular, when CPM is lower (in the model with just EoS), airlines price their fares lower. This is, of course, endogenous. At the same time, when fares are lower (in the model with just EoS), flow increases, possibly causing more links to become capacity constrained,⁵⁶ and driving CPM lower.

⁵⁶I say possibly, because due to substitution between products, it is possible that lower prices on product A may cause consumers to stop purchasing a product B that was previously capacity constrained.

Table 1.38: Counterfactual Variable Profit Margin (%), Full Model

	Predicted	% Change 2012	% Change 2012 (EoS)
Southwest	61	-0.1	-0.1
Delta	45	-2.7	-2.3
United	36	0.3	0.5
American	40	3.3	3.3
US	55	3.7	3.6
JetBlue	52	-0.5	0
AA+US	46	2.5	2.5

In general, profits margins are lower under the full model than under the model with just economies of scale. Under the full model, the airlines face a constrained maximization problem, while in the simpler model, they face an unconstrained maximization problem. That said, there are secondary effects due to the competitive nature of airlines' pricing decisions. I find that these do not overcome the primary effects of constraints: fares are lower in the unconstrained model than the constrained model, for all airlines but US (i.e. AA's acquired products from US). In other words, when facing constraints, airlines are forced to increase average prices on their routes in order to lower demand and prevent it from exceeding capacity.

I find that the merger increases variable profit margins for the merged airline by 2.5%. This is driven partly by its legacy products (3.3% increase) and partly by its acquired products from US (3.7% increase). Because of its increase in costs, Delta is hurt most by the merger: its variable profit margins decrease by 2-3%.

Table 1.39: Counterfactual Summary of Merger: All Carriers, Full Model

	Predicted	% Change 2012	% Change 2012 (EoS)
Median Fares	485	1.2	0.6
Passengers (000s)	2205	-0.6	-0.4
RPM (cents/mile)	11.2	0.9	0.9
CPM (cents/mile)	5.7	0.8	0.7
Variable Profit Margin (%)	98	0.3	0.4
Products (000s)	648	-0.8	-0.8
Consumer Surplus	0	1.5	1.7

Ultimately, a model with capacity constraints can demonstrate the welfare-enhancing effects of the AA-US merger without relying on quality transfer. In particular, I find that the merger increased consumer welfare by 1.5-1.7% *absent* any positive welfare effects of rebranding (i.e. without transferring the AA demand FE to routes acquired from US).

In short, the merger benefits consumers by increasing load factors, causing airlines' costs to go down and allowing them to price lower and attract more passengers. Moreover, it removes a relatively unattractive airline from the industry (US). To expand on this: if, overnight, all US flights had been rebranded as AA flights, we expect consumer welfare to increase. This is driven by our demand estimates: individuals are willing to pay more to travel with AA than US (in particular, 11% more for nonbusiness travelers and 27% more for business travelers). Indeed, if I allow for this quality transfer in the complete model, I find that welfare may increase by over 20% because of the merger. I present these results next.

Quality Transfer

Table 1.40: Counterfactual CPM (cents/mile), Full Model and Quality Transfer

	% Change 2012	% Change 2012 (EoS)
Southwest	−0.2	−0.3
Delta	11.4	3
United	2.5	2.8
American	−6.5	−9.8
US	−30.3	−34.6
JetBlue	−1.3	0
AA+US	−19.5	−23.8

Table 1.41: Counterfactual Capacity Constrained Links, Full Model and Quality Transfer

	% Change 2012	% Change 2012 (EoS)
Southwest	−0.1	−0.1
Delta	−33.8	−12.5
United	−17.5	−20.6
American	30	36.2
US	92.5	92.9
JetBlue	2.4	0.3
AA+US	65.3	68.4

The effects of quality transfer are amplified in a model with endogenous economies of scale. CPM is almost entirely driven by changes in the number of capacity constrained links before/after the merger (the EoS effect). In particular, once AA’s acquired products acquire the AA quality, their demand explodes, and the number of those products that are capacity constrained nearly doubles (93% increase). This causes the costs of AA’s acquired products to decrease by 30-35%. Notice also that Delta loses 12-34% of its capacity constrained links because of the merger — this translates into a cost increase of 3-

11%. On the other hand, AA's legacy products become 30-36% more capacity constrained because of the merger — this translates into a cost decrease of 6-10%. It is likely that the merged airline does not compete much with Southwest, as it barely loses any of its capacity constrained links (0.2%).

Table 1.42: Counterfactual Median Fares, Full Model and Quality Transfer

	% Change 2012	% Change 2012 (EoS)
Southwest	−1	−1.6
Delta	1.8	0.7
United	0	0.6
American	0	−1.1
US	0.4	−1.1
JetBlue	−2.4	−1.5
AA+US	1.6	0.5

Table 1.43: Counterfactual Variable Profit Margin (%), Full Model and Quality Transfer

	% Change 2012	% Change 2012 (EoS)
Southwest	−1	−1.5
Delta	−7.4	−4.4
United	−1.1	−1.3
American	4.8	5.1
US	18.1	18.4
JetBlue	−1.1	−1.4
AA+US	11.2	11.5

There are not many changes in median fares because of the merger.⁵⁷ However, there are large changes in variable profit margins, driven by the decrease in AA's costs as a

⁵⁷The results in the right-most column are not a typo. While AA's median fares decreased on its both its legacy and acquired products (1.1%), overall, AA's median fares increased (0.5%). This is due to the nature of the median function: after the merger, there were more fares on legacy products that fell between the medians of AA's acquired and legacy products. Since the median of its acquired products is lower, this causes the median of legacy products to fall while raising the median of all products, when grouped together.

result of the merger. In particular, I find that the merger increases variable profit margins for the merged airline by 11%. This is driven partly by its legacy products (5% increase in variable profit margins) but mostly by its acquired products from US (18% increase in variable profit margins). Because of its increase in costs, Delta is hurt most by the merger: its variable profit margins decrease by 4-7%.

Table 1.44: Counterfactual Summary of Merger: All Carriers, Full Model and Quality Transfer

	% Change 2012	% Change 2012 (EoS)
Median Fares	0.7	-0.2
Passengers (000s)	9.5	11.6
RPM (cents/mile)	0.4	-1.1
CPM (cents/mile)	-1.1	-2.6
Variable Profit Margin (%)	2.9	3
Products (000s)	-0.8	-0.8
Consumer Surplus	18.2	22.4

Ultimately, a model with capacity constraints can demonstrate the welfare-enhancing effects of the AA-US merger without relying on quality transfer. If I include quality transfer, in addition, I find that the merger increased consumer welfare by 18-22%. This is driven by a drastic reduction in costs of AA's acquired products as they capture economies of scale, caused by their increased demand arising from rebranding and quality transfer.

1.5 Conclusion

In summary, this chapter contributes to the large literature on the estimation of models of airline competition with differentiated products using aggregate data by introducing capacity-constraint considerations on the supply-side, the estimates of which are shown

to be superior to benchmark models in predicting out-of-sample data. I develop a robust approach for ex-post merger evaluation that could be used for future antitrust work and use it to perform a welfare analysis on the AA-US merger.

I find that the merger increased consumer surplus by 1.5-1.7%. Ignoring capacity constraints and economies of scale, the same analysis could have yielded a *decrease* in consumer surplus by as much as 4.5%. The difference between these numbers is driven by two competing factors. On the one hand, due to economies of scale, airlines want to keep flights operating at or close to capacity, so they lower prices of different routes. This lowers marginal costs, and allows them to price even lower, increasing consumer surplus. On the other hand, due to capacity constraints, airlines want to leave some empty room on flights to allow them greater flexibility on pricing the routes in their network. In other words, I find that the AA-US merger benefited consumers and airlines — specifically AA, which enjoyed nearly 2.5% higher profit margins directly as a result of the merger. Benchmark models could lead regulators to erroneously believe this merger harmed consumers.

I also find that there are significant economies of scale in airlines' operations: flights operated at or near capacity had lower marginal costs per passenger by \$128 in 2012 and \$57 in 2015. On the demand-side, I am able to construct a previously unidentified “stopover” variable, and find that the discount consumers require for flights with stopovers is, on average, 3-5 times lower than the discount consumers require for flights with a layover. This suggests that airlines could benefit from turning flights with layovers into flights with stopovers.⁵⁸

⁵⁸Flights with stopovers are still called direct flights and appear as a single flight on an itinerary. This may confuse some passengers, and be responsible for their lower than expected discount, relative to flights with layovers.

A limitation of my model is that the full structural model cannot be estimated: capacity constraints never bind empirically. For this reason, a reduced-form model is used for estimation, while the full model is used for the counterfactuals. Moreover, this paper abstracts from strategic capacity considerations and dynamic pricing: network structure is assumed to be fixed, and pricing decisions, static. I leave such considerations for future work.

Extensions to this paper are discussed more completely in the appendix. One idea would be to include other network constraints, such as pricing constraints. The intuition for this is that airlines are limited in their ability to price routes in their network not only due to capacity considerations, but due to the ability of consumers to “arbitrage” different one-way flights in constructing their itinerary. Other ideas include: making the determinants of marginal cost link instead of product-specific; using additional data from the BTS (the “On-Time Performance” database) to map the routing of individual planes and disaggregate the T-100 capacity data; exploring the effect of multimarket contact in a constrained network; or endogenizing the firm’s choice of capacity. Moreover, one could apply the theory developed here to other observable networks in our economy. I leave this for future work.

2 Efficiency Wages with Health Insurance

2.1 Introduction

The following two chapters are motivated by the hump-shaped lifecycle profile of job search⁵⁹ identified by Aguiar et al. [2013]. Using the American Time Use Survey (“ATUS”), Aguiar et al. [2013] found that this hump-shaped profile persists even after controlling for a host of characteristics, such as education, gender, and family status. This finding was counterintuitive with standard consumption-savings models. In particular, the standard model would predict a decreasing lifecycle profile of job search, driven by the decreasing cost of joblessness over the lifecycle.⁶⁰ Let us call this the **Wealth Effect**.

⁵⁹This is measured by time spent searching for a job.

⁶⁰This is driven by asset accumulation: as individuals age, they become wealthier, and are able to insure themselves against temporary bouts of unemployment while searching for a better job.

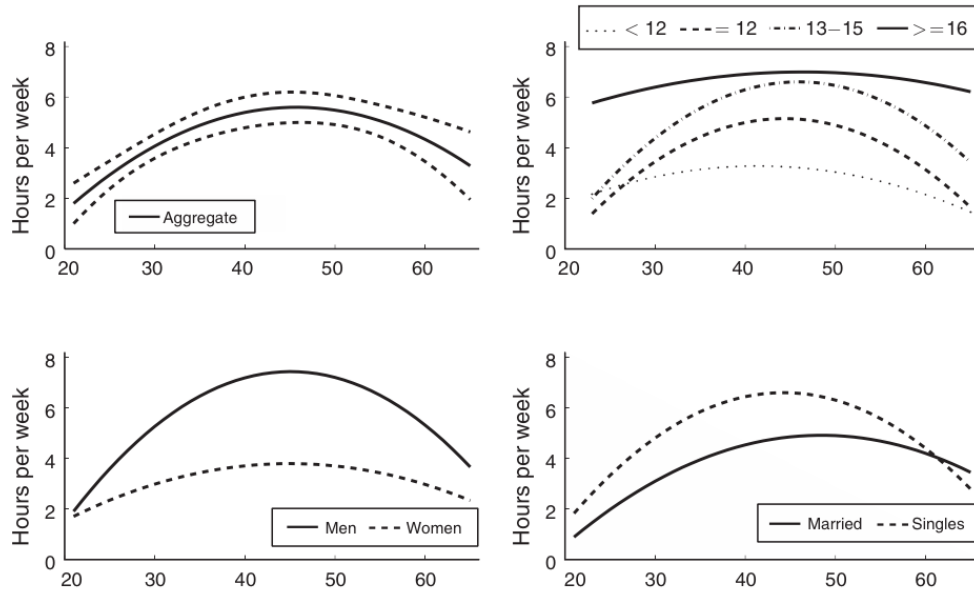


Figure 2.1: Job Search over the Lifecycle - USA (Source: Aguiar et al. [2013])

Moreover, this phenomenon did not extend to other OECD countries. The United Kingdom, Germany, France, and Spain all had the expected decreasing lifecycle profile of job search. There appeared to be some institutional feature specific to the US that caused higher than expected job search in middle age and later life.

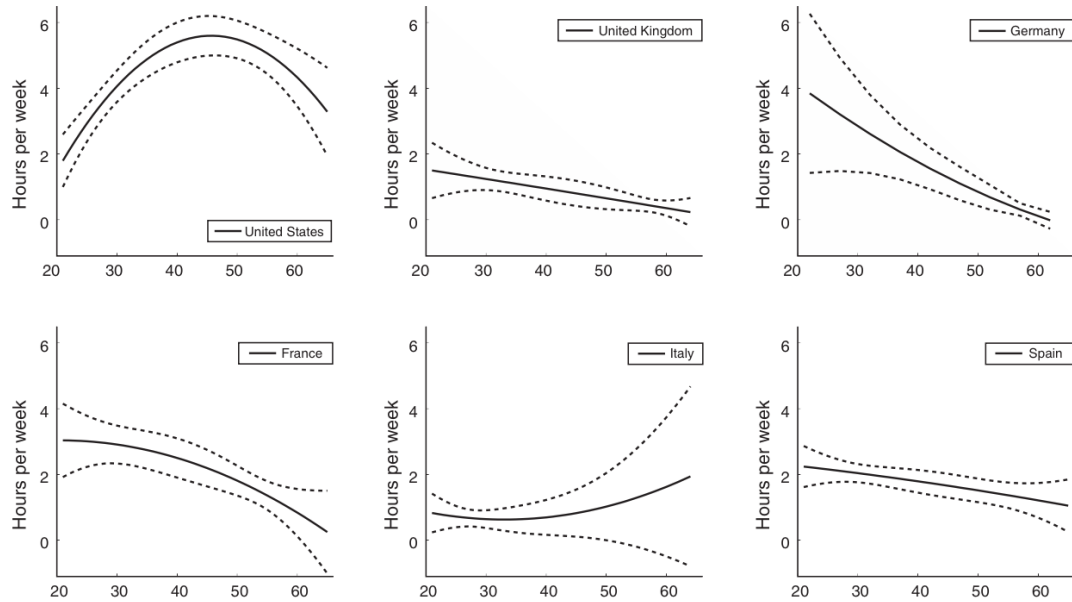
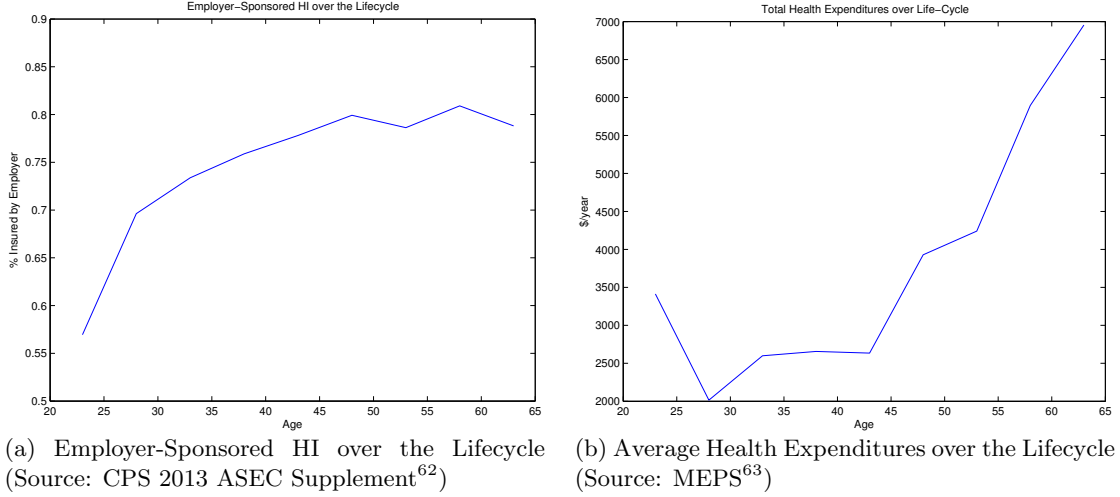


Figure 2.2: Job Search over the Lifecycle - OECD Countries (Source: Aguiar et al. [2013])

One thing that sets the US apart from other OECD countries is its healthcare system, and in particular, its close relationship between health insurance and employment. The majority of Americans are insured by their employer, and this percentage grows over the lifecycle. At the same time, from the mid/late 20's, health expenditures increase over the lifecycle. This lends credence to the following explanation: perhaps the cost of joblessness is higher in middle/old age in the US than in other OECD countries because being jobless often means being uninsured, leading to an *increasing* lifecycle profile of job search as the cost of being uninsured rises over the lifecycle. Let us call this the **Health Effect**.⁶¹

⁶¹Of course, there are other reasons one could think of to explain this phenomenon. For example, one might argue that more experience over the lifecycle increases the opportunity cost of joblessness, raising search effort.

Figure 2.3: Supporting Data from CPS and MEPS



Indeed, with a combination of these two effects (the Wealth Effect and the Health Effect), I am able to generate a lifecycle profile of job search that very closely matches that identified in the ATUS data by Aguiar et al. [2013]. I leave the details of this simple model, the data used, and the simulation results to the appendix.

These findings motivate the next two chapters of this dissertation via the following take-away messages. First, it is clear that employer-sponsored health insurance may play a significant role in providing incentives for employment; this could be dramatically changed by the introduction of the Affordable Care Act (“ACA”). Second, job search is non-constant and endogenous; any theory of optimal taxation must take secondary effects on job search into account. Combining employer-provided health insurance with job search is not novel; Dey and Flinn [2005] do exactly this in their seminal paper.

⁶²In generating a lifecycle profile of job search that very closely matches the ATUS data, I also pooled data from the Current Population Survey (“CPS”) to get unemployment rates and unemployment transitions over many years (as in Choi et al. [2015]).

⁶³To generate this plot, I used data from the 2011-2012, round 16 Medical Expenditure Panel Survey (“MEPS”). In generating a lifecycle profile of job search that very closely matches the ATUS data, I also divided individuals into age bins and self-reported health status. Then, I obtained a distribution of medical health expenditures for these bins, and estimated a health transition matrix using the panel data.

In this chapter, I extend the basic model of efficiency wages (Shapiro and Stiglitz [1984]) to incorporate employer-sponsored health insurance. I show how the ACA reform will affect market wages in a simple economy with imperfect monitoring and a cost-of-effort. Shapiro and Stiglitz [1984] showed that the market equilibrium is not efficient, and that there are circumstances in which the government should intervene in the market by supplying unemployment insurance; I ask whether the same applies to supplying health insurance. To distinguish the individual mandates of the ACA from simply increasing unemployment insurance (now that workers have health insurance even if they get fired, the threat of being fired is lower, akin to increasing unemployment insurance), I introduce worker heterogeneity in the form of health status. Unhealthy workers have higher medical costs and are less productive.

There is a large literature on the role of health insurance on the labor market. See, for example, Currie and Madrian [1999], Li et al. [2013], and Aizawa and Fang [2013]. While Li et al. [2013] attempt to bridge health insurance and efficiency wage theory, they fall short in describing what it is about health insurance that provides utility. I delve inside this “black box” using an empirical framework developed by Aizawa and Fang [2013].

I develop a few theorems, including sufficient conditions under which the Affordable Care Act increases efficiency wages. In particular, if the ACA succeeds, at least in part, in inducing employers to provide health insurance and individuals to self-insure, then wages will rise after its implementation. I suggest that the ACA may provide efficiency wage subsidies towards the socially optimal wage level, and numerically show the existence of regions where this is the case. The intuition for this result is as follows: as subsidies increase, employers are more inclined to offer health insurance to their workers. In turn,

this changes the composition of health status of individuals using the insurance exchanges, driving premiums down. As a result, the relative cost of unemployment goes down, forcing firms to increase wages in order to induce good (no-shirk) behavior.

This chapter is organized as follows. Section 2 outlines the benchmark economic model, designed to capture the environment of firm and worker health insurance decisions prior to the introduction of the ACA. Section 3 introduces the ACA into the benchmark model. Section 4 compares the two, and develops the “Efficiency Wage Theorems”. Section 5 concludes.

2.2 Benchmark Model (Pre-ACA)

2.2.1 Worker’s Problem

Consider an infinite horizon model where individuals choose their level of effort on the job (i.e., whether to shirk, or whether to not shirk), and whether to purchase health insurance if it is not provided to them by their employer or if they are unemployed. An individual can be of either good (**g**) or bad (**b**) health. Let y denote her income, $T(y)$ denote her after-tax income, and x denote her health insurance (“HI”) status, where:

$$x = 0 \Rightarrow \text{worker is uninsured} \tag{2.2.1}$$

$$x = 1 \Rightarrow \text{worker receives HI from her employer} \tag{2.2.2}$$

$$x = 2 \Rightarrow \text{worker purchases HI from the market} \tag{2.2.3}$$

Then, each period, the individual's expected flow utility $v_h(y, x)$ is given by:

$$v_h(y, x) = \begin{cases} \mathbb{E}_{\tilde{m}_h} u(T(y) - \tilde{m}_h), & \text{if } x = 0 \\ u(T(y)), & \text{if } x = 1 \\ u(T(y) - R_h), & \text{if } x = 2 \end{cases} \quad (2.2.4)$$

Where the subscript h denotes the individual's health status (it can take on the values b or g). Notice that in this simple model, health status only affects the individual through her realization of the medical expenditure shock \tilde{m}_h and, as a consequence, health insurance premium R_h . Letting $\bar{m}_h = \mathbb{E}\tilde{m}_h$, and assuming that \tilde{m}_b first order stochastically dominates \tilde{m}_g , I have that $\bar{m}_b > \bar{m}_g$ and that $\mathbb{E}_{\tilde{m}_g} u(T(y) - \tilde{m}_g) > \mathbb{E}_{\tilde{m}_b} u(T(y) - \tilde{m}_b)$. In other words, an unhealthy worker gets, on average, larger medical expenditure shocks than a healthy worker, and workers prefer to be healthy.

Each period, the individual can be either employed or unemployed. While employed, the worker can either exert effort (not shirk) or be lazy (shirk). If the worker does not shirk, she will incur a cost of effort e and get laid off at the exogenous rate of δ . If the worker shirks, on the other hand, she will incur no cost of effort, but may be "discovered" shirking and consequently fired. This happens with probability q , thus increasing her rate of separation to $\delta + q$. Note that q is exogenous in this simple model but can later be made endogenous.

Let $V_h^H(x)$ denote the value function of an employed individual with health status h and health insurance status x , who decides not to shirk (or exert **H**igh effort, H). Similarly, Let $V_h^L(x)$ denote the value function of an employed individual with health status h and

health insurance status x , who decides to shirk (or exert **L**ow effort, **L**). I can write the value functions as:

$$rV_h^H(x) = v_h(y, x) - e + \delta(U_h(\hat{x}_h) - \max\{V_h^H(x), V_h^L(x)\}) \quad (2.2.5)$$

$$rV_h^L(x) = v_h(y, x) + (\delta + q)(U_h(\hat{x}_h) - \max\{V_h^H(x), V_h^L(x)\}) \quad (2.2.6)$$

Now, $U_h(x)$ is the value function of an unemployed individual with health status h and health insurance status x (to be defined explicitly below), and \hat{x}_h is the individual's health insurance status immediately after being laid off (also to be defined explicitly below). In the US, under the Consolidated Omnibus Budget Reconciliation Act of 1985 (“COBRA”), employers are strongly encouraged (via tax deductions) to extend employer-provided health insurance for a period of time after an involuntary separation (provided “gross misconduct” did not occur). Since separations in my model are involuntary, I extend employer-provided health insurance for a single period after an exogenous separation. Notice, however, that if the worker chose to self-insure herself while employed, she is not required to continue with that coverage. Indeed, it may be optimal for her to terminate her coverage (or begin covering herself, if she previously had not), given that her income will have changed from y to b .⁶⁴

In other words, an individual's health insurance status immediately after being laid off is given by:

$$\hat{x}_h = \begin{cases} x, & \text{if } x = 1 \\ x_h^*, & \text{otherwise} \end{cases} \quad (2.2.7)$$

⁶⁴Possible extension: if discovered shirking, consider this a “gross misconduct” and prevent COBRA continuation payments in that case.

Where $x_h^* = \arg \max_{x' \in \{0,2\}} U_h(x') = \arg \max_{x' \in \{0,2\}} v_h(b, x')$. Next, let us write the value function of an unemployed individual with health status h and health insurance status $x \in \{0, 2\}$:

$$(1 + r)U_h(x) = v_h(b, x) + \alpha \max\{V_h^H(y, \hat{x}_h), V_h^L(y, \hat{x}_h)\} + (1 - \alpha)U_h(x_h^*) \quad (2.2.8)$$

Where

$$\hat{x}_h = \begin{cases} 1, & \text{if HI offered} \\ x_h^{**}, & \text{otherwise} \end{cases} \quad (2.2.9)$$

And x_h^{**} , the optimal choice of whether to purchase HI while employed when the employer does not offer HI, satisfies: $x_h^{**} = \arg \max_{x' \in \{0,2\}} v_h(y, x')$.⁶⁵

Notice that the individual will always accept the (y, x) offer, provided $y > b$, since she can always shirk on the job and earn the wage of y at no additional cost to herself. In other words, being employed and shirking is strictly better than being unemployed. The workers only choices, then, are:

- 1) x_h^* : whether to purchase HI when employed, if employers do not offer HI.
- 2) x_h^{**} : whether to purchase HI when unemployed, after COBRA continuation ends, or if employers did not provide HI.

Since the individual's problem is stationary (health status does not change), if the worker shirked today, she will necessarily shirk again tomorrow (the same goes for exerting effort). In other words, "once a shirker always a shirker". I can get rid of the max operator

⁶⁵This follows from equations (2.2.12) and (2.2.13). Note that x^* may not equal x^{**} , since the flow utility $v_h(y, x)$ at $x = 2$ depends on income, which differs when employed (y) or unemployed (b).

in the value functions while employed:

$$rV_h^H(x) = v_h(y, x) - e + \delta(U_h(\hat{x}_h) - V_h^H(x)) \quad (2.2.10)$$

$$rV_h^L(x) = v_h(y, x) + (\delta + q)(U_h(\hat{x}_h) - V_h^L(x)) \quad (2.2.11)$$

Solving, I get:

$$V_h^H(x) = \frac{v_h(y, x) - e + \delta U_h(\hat{x}_h)}{r + \delta} \quad (2.2.12)$$

$$V_h^L(x) = \frac{v_h(y, x) + (\delta + q)U_h(\hat{x}_h)}{r + \delta + q} \quad (2.2.13)$$

Then, I have two No-Shirk-Conditions (“NSC’s”): $V_h^H(\hat{x}_h) > V_h^L(\hat{x}_h)$ for $h = b, g$.

From the previous value function expressions, I can re-write the NSC’s as:

$$v_h(y, \hat{x}_h) > rU_h(\hat{x}_h) + (r + \delta + q)e/q \equiv \hat{v}_h(\hat{x}_h) \quad (2.2.14)$$

Notice immediately that offering health insurance is a “double-edged sword” (recall that if the firm offers HI, $\hat{x}_h = \hat{x}_h = 1$). On one hand, it raises $v_h(y, x)$ and adds to the efficiency wage compensation required to get workers not to shirk, but on the other hand it increases the value of unemployment for workers, since after termination they get to continue with their health insurance coverage for a few months,⁶⁶ raising $\hat{v}_h(\hat{x}_h)$.

⁶⁶In my model, this continues until the next period, which is defined as the average length of COBRA continuation.

2.2.2 Health Insurance Purchase Decision

If individuals choose to purchase health insurance (either while unemployed after the completion of COBRA continuation, or while employed if their employer does not offer HI), they will be charged a premium proportional to their average medical expenditure, given by:

$$R_h = (1 + \xi_h)\bar{m}_h \quad (2.2.15)$$

Where \bar{m}_h is the average medical expenditure for individuals of type h health status, and $\xi_h > 0$ is a loading factor for the health insurance market, which I allow to be health-type specific. I bound ξ_h by assuming that $R_g > R_b$.

Let us characterize the individual's Purchase/No Purchase HI decision. Doing so requires assumptions on the function form of the utility function. Consider first the case when my utility function exhibits constant absolute risk aversion ("CARA"),⁶⁷ i.e., $u(x) = -e^{-\gamma x}$. Then, if I assume that medical expenditure shocks are normally distributed, I can express the certainty equivalent of the no-HI gamble by:⁶⁸

$$CE(x = 0) = \mu - \frac{1}{2}\gamma\sigma^2 \quad (2.2.16)$$

Where, in my example, $\mu = T(y) - \bar{m}_h$ and $\sigma^2 = Var(\tilde{m}_h)$. In other words:

$$\mathbb{E}_{\tilde{m}_h} u(T(y) - \tilde{m}_h) = u(T(y) - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h)) \quad (2.2.17)$$

⁶⁷This allows me to handle negative consumption in the case of a particularly bad medical expenditure shock.

⁶⁸For a complete derivation, refer to the appendix, section B.2.1.

Under this functional form assumption (CARA utility), I can find necessary and sufficient conditions for when an individual elects to purchase health insurance from the market:

Theorem 1. *If u exhibits CARA,*

$$x_h^* = 2 \iff x_h^{**} = 2 \iff R_h < \frac{1}{\gamma} \log \int e^{\gamma x} f_h(x) dx \iff \xi_h < \frac{1}{\gamma \bar{m}_h} \log \int e^{\gamma x} f_h(x) dx - 1$$

Moreover, if medical expenditure shocks are normally distributed,

$$x_h^* = 2 \iff x_h^{**} = 2 \iff R_h < \bar{m}_h + \frac{1}{2} \gamma \text{Var}(\tilde{m}_h) \iff \xi_h < \frac{1}{2} \gamma \frac{\text{Var}(\tilde{m}_h)}{\bar{m}_h}$$

Proof. In appendix. □

Next, I relax the assumption of constant absolute risk aversion to the class of utility functions that exhibit decreasing absolute risk aversion (“DARA”). I can find sufficient conditions for when an employed individual not getting health insurance from his employer elects to purchase health insurance from the market:

Theorem 2. *If u exhibits DARA, $x_h^{**} = 2 \Rightarrow x_h^* = 2$*

Proof. In appendix. □

As a corollary, I can also find sufficient conditions for when an unemployed individual elects to remain uninsured:

Corollary 1. *If u exhibits DARA, $x_h^* = 0 \Rightarrow x_h^{**} = 0$*

Finally, I can find analogs of the sufficient conditions above for the special case of utility exhibiting increasing absolute risk aversion (“IARA”):

Corollary 2. *If u exhibits IARA, $x_h^* = 2 \Rightarrow x_h^{**} = 2$ and $x_h^{**} = 0 \Rightarrow x_h^* = 0$*

2.2.3 Employer's Problem

The representative firm gets nothing if it hires a worker who shirks, p units of labor if it hires a healthy worker who does not shirk, and pd units of labor if it hires an unhealthy worker who does not shirk, where $d < 1$ denotes the loss of productivity due to bad health (this includes time spent not working due to sickness, as well as lower productivity while on the job).

If the firm offers HI to its workers, it incurs a variable cost equal to those workers' average medical expenditures, \bar{m}_h , in addition to a fixed administrative cost of offering health insurance, C . This means that the representative firm has 6 possible strategies (in addition to a shutdown option):

Strategy 1: Do not offer HI, set wage to ensure NSC holds for workers of both good and bad health, y_{all}^0 , and earn corresponding profits equal to π_{all}^0 .

Strategies 2 and 3: Do not offer HI, set wage to ensure NSC holds for only one type of worker, y_b^0 **or** y_g^0 , and earn corresponding profits equal to π_b^0 **or** π_g^0 , respectively.

Strategy 4: Offer HI, set wage to ensure NSC holds for workers of both good and bad health, y_{all}^1 , and earn corresponding profits equal to π_{all}^1 .

Strategies 5 and 6: Offer HI, set wage to ensure NSC holds for only one type of worker, y_b^1 **or** y_g^1 , and earn corresponding profits equal to π_b^1 **or** π_g^1 , respectively.

Then, the firm solves:

$$\max\{\pi_{all}^0, \pi_b^0, \pi_g^0, \pi_{all}^1, \pi_b^1, \pi_g^1, 0\} \quad (2.2.18)$$

where

$$\begin{aligned}
\pi_{all}^0 &= (pd - y_{all}^0)n_b(y_{all}^0, 0) + (p - y_{all}^0)n_g(y_{all}^0, 0) \\
\pi_b^0 &= (pd - y_b^0)n_b(y_b^0, 0) + (0 - y_b^0)n_g(y_b^0, 0) \\
\pi_g^0 &= (0 - y_g^0)n_b(y_g^0, 0) + (p - y_g^0)n_g(y_g^0, 0) \\
\pi_{all}^1 &= (pd - y_{all}^1 - \bar{m}_b)n_b(y_{all}^1, 1) + (p - y_{all}^1 - \bar{m}_g)n_g(y_{all}^1, 1) - C \\
\pi_b^1 &= (pd - y_b^1 - \bar{m}_b)n_b(y_b^1, 1) + (0 - y_b^1 - \bar{m}_g)n_g(y_b^1, 1) - C \\
\pi_g^1 &= (0 - y_g^1 - \bar{m}_b)n_b(y_g^1, 1) + (p - y_g^1 - \bar{m}_g)n_g(y_g^1, 1) - C
\end{aligned}$$

where $n_h(y, x)$ is the steady-state level of employment of workers of health status h at the wage and health insurance offer (y, x) . Notice that $(y_{all}^0, y_b^0, y_g^0, y_{all}^1, y_b^1, y_g^1)$ are given implicitly by the equations:

$$v_h(y_h^x, \hat{x}_h) = rU_h(\hat{x}_h) + (r + \delta + q)e/q \quad (2.2.19)$$

$$y_{all}^x = \max\{y_b^x, y_g^x\} \quad (2.2.20)$$

Next, notice that the steady-state levels of employment are very easy to derive. Given G workers of good health and B workers of bad health, the steady-state levels of employment are given by:

$$n_g(y_g^0, 0) = n_g(y_g^1, 1) = n_g(y_{all}^0, 0) = n_g(y_{all}^1, 1) = \frac{\alpha G}{\alpha + \delta} \quad (2.2.21)$$

$$n_b(y_b^0, 0) = n_b(y_b^1, 1) = n_b(y_{all}^0, 0) = n_b(y_{all}^1, 1) = \frac{\alpha B}{\alpha + \delta} \quad (2.2.22)$$

$$n_g(y_b^0, 0) = n_g(y_b^1, 1) = \frac{\alpha G}{\alpha + \delta + q} \quad (2.2.23)$$

$$n_b(y_g^0, 0) = n_b(y_g^1, 1) = \frac{\alpha B}{\alpha + \delta + q} \quad (2.2.24)$$

This follows from the fact that at y_g^0 and y_g^1 , all good health workers are taking every job that comes to them and not shirking, while at y_b^0 and y_b^1 , they are still taking every job that comes to them **but** shirking. In the steady state, the flow in must equal the flow out, so $\alpha(G - n_g(y_g)) = \delta n_g(y_g) \Rightarrow n_g(y_g) = \frac{\alpha G}{\alpha + \delta}$, and $\alpha(G - n_g(y_b)) = (\delta + q)n_g(y_b) \Rightarrow n_g(y_b) = \frac{\alpha G}{\alpha + \delta + q}$. The same exercise can be repeated for the workers of bad health status.

2.2.4 Solving for No-Shirk-Conditions

It will be convenient to re-write the value function of unemployment in a way that takes advantage of the nature of COBRA continuation and the static benefit of health insurance in my model. Since COBRA continuation only extends employer-sponsored health insurance for one period after lay-off, and since the benefits of health insurance come about only from changing v_h , I can re-write the value function of unemployment for a type h health status worker facing the equilibrium wage offer y as:⁶⁹

⁶⁹I will omit specifying x explicitly, since I can write any y as y_i^j where i captures whose no-shirk constraints are satisfied, and j captures whether health insurance is offered or not, in order to minimize notation.

$$U_h^y(x) = \frac{1}{1+r} [v_h(b, x) + U_h^y] \quad (2.2.25)$$

where U_h^y does not depend on x :

$$U_h^y = (1 - \alpha)[U_h^y(x_h^*)] + \alpha \begin{cases} V_h^H(y, 1), & \text{if HI offered and NSC holds} \\ V_h^L(y, 1), & \text{if HI offered and NSC does not hold} \\ V_h^H(y, x^{**}), & \text{if HI not offered and NSC holds} \\ V_h^L(y, x^{**}), & \text{if HI not offered and NSC does not hold} \end{cases}$$

Solving explicitly (derivation is in the appendix), I get:

$$U_h^y = \begin{cases} \frac{(1+r)(r+\delta)(\frac{1-\alpha}{\alpha})v_h(b, x^*) + (1+r)(v_h(y, 1) - e) + \delta v_h(b, 1)}{r(1+\delta+r)}, & \text{if HI and NSC} \\ \frac{(1+r)(r+\delta+q)(\frac{1-\alpha}{\alpha})v_h(b, x^*) + (1+r)(v_h(y, 1)) + (\delta+q)v_h(b, 1)}{r(1+\delta+r+q)}, & \text{if HI and no NSC} \\ \frac{(1+r)(r+\delta)(\frac{1-\alpha}{\alpha})v_h(b, x^*) + (1+r)(v_h(y, x^{**}) - e) + \delta v_h(b, x^*)}{r(1+\delta+r)}, & \text{if no HI but NSC} \\ \frac{(1+r)(r+\delta+q)(\frac{1-\alpha}{\alpha})v_h(b, x^*) + (1+r)(v_h(y, x^{**}) - e) + (\delta+q)v_h(b, x^*)}{r(1+\delta+r+q)}, & \text{if no HI and no NSC} \end{cases}$$

Notice that since I am defining separate value functions of unemployment based on “Shirk Status”, my no-shirk condition will need to be slightly changed to reflect this. In particular I must treat the U_h in the value function for a non-shirker differently from the U_h in the value function for a shirker, as written above.

Case 1: Firm offers HI

In particular, the no-shirk condition in the world where the representative firm offers health insurance pins down the equilibrium y :

$$\begin{aligned} V_h^H(1) &= V_h^L(1) \\ \frac{v_h(y, 1) - e + \delta U_h^{y_h^1}(1)}{r + \delta} &= \frac{v_h(y, 1) + (\delta + q)U_h^{y_h^1}(1)}{r + \delta + q} \end{aligned}$$

Since $U_h^{y_h^1}(1) = \frac{1}{1+r}[v_h(b, 1) + U_h^y]$, I can solve explicitly for $v(y, 1)$ to get (full derivation in appendix):

$$\Rightarrow v(y, 1) = v(b, 1) + \left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1 + \delta + q + r)e/q \quad (2.2.26)$$

In other words, the firm will set y such that the worker is indifferent between shirking and not shirking (in which case he will not shirk). Notice that since the firm offers health insurance in equilibrium in this case, the wage y it must offer to achieve no-shirking is higher—since $v(b, 1) > v(b, 0)$ and $v(b, 1) > v(b, 2)$. This again illustrates the “double-edged sword” nature of offering in health insurance.

Next, notice that since only $v(b, x^*)$ depends on health status in the above equality (since the firm offers health insurance), and $v_g(b, x^*) \geq v_b(b, x^*)$, it is necessarily the case that the NSC of the good-health status worker will be binding: $y_g^1 \geq y_b^1$. This motivates the following theorem:

Theorem 3. *When the firm offers health insurance, the NSC of the good-health status worker will be binding: $y_{all}^1 = y_g^1 \geq y_b^1$.*

Proof. By contradiction. Consider if $y_b^1 > y_g^1$. Then at y_b^1 , both NSC conditions are satisfied. In particular, the NSC is satisfied with equality for the bad-health status worker. Since $R_g \leq R_b$, I have that $v_g(b, x^*) \geq v_b(b, x^*)$, so the right-hand side is at least as large for the good-health status worker (while the left-hand side is independent of health status). Then, if $v_g(b, x^*) > v_b(b, x^*)$, the good-health status worker would find it optimal to shirk. On the other hand, if $v_g(b, x^*) = v_b(b, x^*)$, the right-hand side would also be independent of health status, so I must have that $y_b^1 = y_g^1$. Both cases lead to the desired contradiction. \square

Case 2: Firm does not offer HI

Now, the NSC is determined by:

$$\begin{aligned} V_h^H(0) &= V_h^L(0) \\ \frac{v_h(y, x^{**}) - e + \delta U_h^{y_h^0}(x^*)}{r + \delta} &= \frac{v_h(y, x^{**}) + (\delta + q)U_h^{y_h^0}(x^*)}{r + \delta + q} \end{aligned}$$

Since $U_h^{y_h^0}(x^*) = \frac{1}{1+r}[v_h(b, x^*) + U_h^y]$, I can solve explicitly for $v(y, x^{**})$ to get (full derivation in appendix):

$$\Rightarrow v_h(y, x^{**}) = \frac{1}{\alpha} v_h(b, x^*) + (1 + \delta + q + r)e/q \quad (2.2.27)$$

In other words, the firm will set y such that the worker is indifferent between shirking and not shirking (in which case he will not shirk).

I next present the analog of Theorem 3 for the case where the firm does not offer HI:

Theorem 4. *When the firm does not offer health insurance and u exhibits CARA, the NSC of the good-health status worker will be binding: $y_{all}^0 = y_g^0 > y_b^0$.*

Proof. In appendix. □

Combining Cases:

Next, I compare the NSC condition when the firm offers HI to when the firm does not offer HI. In other words, I compare (2.2.26) to (2.2.27).

From (2.2.26), I have that:

$$\begin{aligned} v(y, 1) &= v(b, 1) + \left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1 + \delta + q + r)e/q \\ &= \frac{1}{\alpha}v_h(b, x^*) + [v(b, 1) - v_h(b, x^*)] + (1 + \delta + q + r)e/q \end{aligned}$$

Subtracting (2.2.27) from this expression, I get:

$$\begin{aligned} v(y^1, 1) - v_h(y^0, x^{**}) &= v(b, 1) - v_h(b, x^*) \tag{2.2.28} \\ \Rightarrow u(y^1) - \max\{u(y^0 - R_h), \mathbb{E}u(y^0 - \tilde{m}_h)\} &= u(b) - \max\{u(b - R_h), \mathbb{E}u(b - \tilde{m}_h)\} \end{aligned}$$

Theorem 5. *When u exhibits CARA, $y^1 > y^0$.*

Proof. In appendix. □

As before, let us relax the CARA assumption to the class of DARA utility functions:

Theorem 6. *When u exhibits DARA, $y^1 > y^0$.*

Proof. In appendix. □

Let us provide some intuition for this counter-intuitive result. Notice that I can write the two NSC conditions as follows:

$$\begin{aligned} v(y^1, 1) &= \frac{1}{\alpha} v(b, x^*) + K + v(b, 1) - v(b, x^*) \\ v(y^0, x^{**}) &= \frac{1}{\alpha} v(b, x^*) + K \end{aligned} \tag{2.2.29}$$

When the firm offer health insurance, it *increases the benefit of shirking* by $v(b, 1) - v(b, x^*)$, since now if the worker shirks, he will get free health insurance while unemployed, as opposed to paying for it or being uninsured (x^*). At the same time, though, it *increases the benefit of not shirking* by $v(y, 1) - v(y, x^{**})$, since now if the worker does not shirk, he is more likely to stay employed and get health insurance, instead of paying for it or remaining uninsured (x^{**}). This is the trade-off the firm must evaluate when it decides whether or not to offer health insurance. Absent subsidies or penalties, though, since utility is concave, additional consumption at lower wealth is *more valuable*. Since $x^* \iff x^{**}$, the individual is getting the same “additional unit of consumption” by shirking and by not shirking, but this additional unit is more valuable at his lower wealth level, b , increasing the motivation to shirk. To compensate for this phenomenon, the firm must offer a greater wage compensation package when it offers health insurance. I am now ready to present my impossibility theorem for the benchmark model:

Theorem 7 (Impossibility Result). *If u exhibits CARA or DARA, the firm will never find it optimal to offer health insurance (absent government intervention).*

Proof. Follows from Theorems 5 and 6. Since $y^1 > y^0$, given y , the firm attracts *no more* workers by offering health insurance, and offering health insurance is costly. In other

words, there is no reason to offer health insurance in my benchmark model, given CARA or DARA utility. \square

I note that if u exhibits IARA, the firm may find it optimal to offer health insurance absent government intervention, since now the lower value of an additional unit of consumption at higher wealth is offset by the increased aversion of risk at that higher wealth. However, there is little empirical support for IARA in the real-world.⁷⁰

2.2.5 Solving for Equilibrium Wage

Knowing this, I can easily solve for the equilibrium wage directly. From theorem 4, I have that if u exhibits CARA, the NSC of the good-health status worker will be binding. If this is the case, then the firm has two options:

- 1) Set $y = y_g^0$, earn $\pi = (pd - y)\frac{\alpha B}{\alpha + \delta} + (p - y)\frac{\alpha G}{\alpha + \delta}$
- 2) Set $y = y_b^0$, earn $\pi = (pd - y)\frac{\alpha B}{\alpha + \delta} + (0 - y)\frac{\alpha G}{\alpha + \delta + q}$

Moreover, I can solve for y_g^0 and y_b^0 via $v(y, x^{**}) = \frac{1}{\alpha}v(b, x^*) + (1 + \delta + q + r)e/q$.

If u exhibits CARA, theorem 2.1 tells us that $x_h^* = 2 \iff x_h^{**} = 2$. This leaves us with only 2 cases to consider:

⁷⁰Other ways to remedy this impossibility result are by: 1) stopping COBRA continuation after an employee is caught shirking, or 2) introducing a direct benefit of HI to employee productivity.

Case 1: $x_h^* = x_h^{**} = 2$. Then, letting $(1 + \delta + q + r)e/q = Q$, the NSC reduces to:

$$\begin{aligned}
v(y, 2) &= \frac{1}{\alpha}v(b, 2) + (1 + \delta + q + r)e/q \\
u(T(y) - (1 + \xi_h)\bar{m}_h) &= \frac{1}{\alpha}u(T(b) - (1 + \xi_h)\bar{m}_h) + Q \\
-\alpha e^{-\gamma[T(y) - (1 + \xi_h)\bar{m}_h]} &= -e^{-\gamma[T(b) - (1 + \xi_h)\bar{m}_h]} + \alpha Q \\
\alpha e^{-\gamma T(y)} &= e^{-\gamma T(b)} - \alpha Q e^{-\gamma(1 + \xi_h)\bar{m}_h} \\
\log(\alpha) - \gamma T(y) &= \log(e^{-\gamma T(b)} - \alpha Q e^{-\gamma(1 + \xi_h)\bar{m}_h}) \\
T(y_h^0) &= \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma T(b)} - \alpha Q e^{-\gamma(1 + \xi_h)\bar{m}_h}}\right)
\end{aligned}$$

Case 2: $x_h^* = x_h^{**} = 0$. Then, letting $(1 + \delta + q + r)e/q = Q$, the NSC reduces to:

$$\begin{aligned}
v(y, 0) &= \frac{1}{\alpha}v(b, 0) + (1 + \delta + q + r)e/q \\
u(T(y) - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h)) &= \frac{1}{\alpha}u(T(b) - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h)) + Q \\
-\alpha e^{-\gamma[T(y) - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h)]} &= -e^{-\gamma[T(b) - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h)]} + \alpha Q \\
\alpha e^{-\gamma T(y)} &= e^{-\gamma T(b)} - \alpha Q e^{-\gamma[\bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h)]} \\
\log(\alpha) - \gamma T(y) &= \log(e^{-\gamma T(b)} - \alpha Q e^{-\gamma[\bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h)]}) \\
T(y_h^0) &= \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma T(b)} - \alpha Q e^{-\gamma[\bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h)]}}\right)
\end{aligned}$$

It follows then that if $e^{-\gamma[\bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h)]} > e^{-\gamma(1 + \xi_h)\bar{m}_h}$, that $y|(x_h^* = 0) > y|(x_h^* = 2)$.

Or, if $\bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h) < (1 + \xi_h)\bar{m}_h$. But notice that this is exactly the condition I had

to ensure that $x_h^* = x_h^{**} = 0$.

Then, I can evaluate profits from the firm's two options to determine whether the firm will only ensure that the NSC of the bad-health status worker will hold, or whether it will ensure that both hold. The firm will ensure that all NSC's hold if and only if $p \frac{\alpha G}{\alpha + \delta} > (y_g - y_b) \frac{\alpha B}{\alpha + \delta} + y_g \frac{\alpha G}{\alpha + \delta} - y_b \frac{\alpha B}{\alpha + \delta + q}$. The results above completely characterize the benchmark model for CARA utility. Simulation can be used to characterize the model for other utility functions, like DARA utility.

2.3 Affordable Care Act

In early 2010, President Barack Obama signed the Patient Protection and Affordable Care Act (PPACA), commonly called the Affordable Care Act (ACA) or "Obamacare" into law. Amongst other things, the ACA stipulated that there be:

Guaranteed issue, which prohibits insurers from denying coverage to individuals due to pre-existing conditions, and a partial community rating requires insurers to offer the same premium price to all applicants of the same age and geographical location without regard to gender or most pre-existing conditions.

Health insurance exchanges that commence operation in every state. Each exchange will serve as an online marketplace where individuals and small businesses can compare policies and buy insurance (with a government subsidy if eligible).

An **individual mandate**, which requires all individuals not covered by an employer sponsored health plan, Medicaid, Medicare or other public insurance programs (such as Tricare) to secure an approved private-insurance policy or pay a penalty.

An **employer mandate**, which requires that businesses which employ 50 or more

people but do not offer health insurance to their full-time employees to pay a tax penalty if the government has subsidized a full-time employee's healthcare through tax deductions or other means.

In this paper, I model the **guaranteed issue** and **health insurance exchanges** provisions by introducing a health insurance exchange that can only charge a single insurance premium, R^{EX} , and cannot deny coverage, to both good and bad health individuals. I allow the potential government subsidy to depend on both income and R^{EX} and denote it by $S(y, R^{EX})$.

I model the **individual mandate** by introducing a penalty when health insurance is not purchased, as a function of income: $P_W(y)$, where the subscript W denotes "worker".

I model the **employer mandate** by introducing a penalty when health insurance is not offered by the employer, as a function of the number of employed workers: $P_E(n)$, where the subscript E denotes "employer".

In other words, the world under the ACA differs from the pre-ACA, benchmark model in the following ways:

1) Existence of Insurance Exchange: insurance rates will be the same for good and bad health individuals, since discrimination on medical history and preexisting conditions is not allowed.

2) Implementation of Penalties and Subsidies: for both individuals and employers.

I now proceed to characterize the ACA model.

2.3.1 Worker's Problem

Following Aizawa and Fang [2013], the expected flow utility $v_h(y, x)$ can be re-written as:

$$v_h(y, x) = \begin{cases} \mathbb{E}_{\tilde{m}_h} u(T(y) - \tilde{m}_h - P_W(y)), & \text{if } x = 0 \\ u(T(y)), & \text{if } x = 1 \\ u(T(y) + S(y, R^{EX}) - R^{EX}), & \text{if } x = 2 \end{cases} \quad (2.3.1)$$

Recall Theorem (1). Analogously, in the world where the ACA exists, I get the following theorem:

Theorem 8. *If u exhibits CARA,*

$$\begin{aligned} x_h^* = 2 & \iff R^{EX} < \frac{1}{\gamma} \log \int e^{\gamma x} f_h(x) dx + S(b) + P(b) \\ x_h^{**} = 2 & \iff R^{EX} < \frac{1}{\gamma} \log \int e^{\gamma x} f_h(x) dx + S(y) + P(y) \end{aligned}$$

Moreover, if medical expenditure shocks are normally distributed,

$$x_h^* = 2 \iff R^{EX} < S(b, R^{EX}) + \bar{m}_h + \frac{1}{2}\gamma \text{Var}(\tilde{m}_h) + P_W(b) \quad (2.3.2)$$

$$x_h^{**} = 2 \iff R^{EX} < S(y, R^{EX}) + \bar{m}_h + \frac{1}{2}\gamma \text{Var}(\tilde{m}_h) + P_W(y) \quad (2.3.3)$$

Furthermore, if $S(y, R^{EX})$ and $P_W(y)$ take the functional forms:

$$P_W(y) = k_1 y$$

$$S(y, R^{EX}) = k_2 R^{EX} - k_3 y$$

And $S(y, R^{EX}) \geq 0$, then:

$$k_1 > k_3 \Rightarrow (x_h^* = 2 \Rightarrow x_h^{**} = 2) \quad (2.3.4)$$

$$k_1 < k_3 \Rightarrow (x_h^* = 2 \Leftarrow x_h^{**} = 2) \quad (2.3.5)$$

$$k_1 = k_3 \Rightarrow (x_h^* = 2 \Longleftrightarrow x_h^{**} = 2) \quad (2.3.6)$$

Proof. If u is CARA, then it must necessarily be of the exponential form $u(x) = -e^{-\gamma x}$.

Then,

$$\begin{aligned} x_h^{**} = 2 &\Longleftrightarrow -e^{-\gamma[T(y)-R+S(y)]} > \int -e^{-\gamma[T(y)-P(y)-x]} f(x) dx \\ &\Longleftrightarrow -e^{-\gamma[T(y)-R+S(y)]} > -e^{-\gamma[T(y)-P(y)]} \int e^{\gamma x} f(x) dx \\ &\Longleftrightarrow e^{-\gamma[T(y)-R+S(y)]+\gamma[T(y)-P(y)]} < \int e^{\gamma x} f(x) dx \\ &\Longleftrightarrow e^{\gamma[R-S(y)-P(y)]} < \int e^{\gamma x} f(x) dx \\ &\Longleftrightarrow R < \frac{1}{\gamma} \log \int e^{\gamma x} f(x) dx + S(y) + P(y) \end{aligned}$$

Similarly,

$$x_h^* = 2 \Longleftrightarrow R < \frac{1}{\gamma} \log \int e^{\gamma x} f(x) dx + S(b) + P(b)$$

The second part follows from: $\mathbb{E}_{\tilde{m}_h} u(T(y) - \tilde{m}_h - P_W(y)) = u(T(y) - \tilde{m}_h - P_W(y) - \frac{1}{2}\gamma Var(\tilde{m}_h))$. The third part follows from noticing that the right-hand side of the first two if and only if expressions above changes by $-k_3(y - b) + k_1(y - b)$ as we move from the unemployed to employed world. If $k_1 \geq k_3$, this net change is non-negative, so if the individual purchased health insurance while unemployed, he will surely want to purchase

it now that he is employed. The converse holds if $k_1 \leq k_3$. □

2.3.2 Employer's Problem

The employer's problem remains unchanged except for the imposition of a penalty term if the firm does not offer health insurance to its workers, which is a function of the size of the firm: $P_E(n)$. Profits when the firm does not offer health insurance change to the following:

$$\pi_{all}^0 = (pd - y_{all}^0)n_b(y_{all}^0, 0) + (p - y_{all}^0)n_g(y_{all}^0, 0) - P_E(n(y_{all}^0, 0)) \quad (2.3.7)$$

$$\pi_b^0 = (pd - y_b^0)n_b(y_b^0, 0) + (0 - y_b^0)n_g(y_b^0, 0) - P_E(n(y_b^0, 0)) \quad (2.3.8)$$

$$\pi_g^0 = (0 - y_g^0)n_b(y_g^0, 0) + (p - y_g^0)n_g(y_g^0, 0) - P_E(n(y_g^0, 0)) \quad (2.3.9)$$

Where $n(y, 0) = n_g(y, 0) + n_b(y, 0)$. Recall that the employer's choice of wage offer, y , is such that some of the NSC conditions just bind:

$$v_h(y_h^x, \hat{x}_h) = rU_h(\hat{x}_h) + (r + \delta + q)e/q \quad (2.3.10)$$

$$y_{all}^x = \max\{y_b^x, y_g^x\} \quad (2.3.11)$$

Explicitly,

$$v_h(y_h^1, 1) = u(T(y_h^1)) = rU_h(1) + (r + \delta + q)e/q \quad (2.3.12)$$

$$v_h(y_h^0, x_h^{**}) = rU_h(x_h^*) + (r + \delta + q)e/q \quad (2.3.13)$$

$$x_h^{**} = \arg \max_{x' \in \{0,2\}} v_h(y_h^0, x') \quad (2.3.14)$$

$$x_h^* = \arg \max_{x' \in \{0,2\}} v_h(b, x') \quad (2.3.15)$$

Now, $v_b(y, 2) = v_g(y, 2) \quad \forall y$, since the health insurance exchange cannot discriminate on pre-existing conditions (i.e., bad health). This implies that when the employer offer health insurance, if both good and bad health workers choose to insure themselves when unemployed, i.e., $x^* = 2$, I get that $U_g(1) = U_b(1)$, so $y_b^1 = y_g^1 = y_{all}^1 = y^1$.

Similarly, if the employer does not offer health insurance, and both good and bad health workers choose to insure themselves when unemployed *and* when employed, i.e., $x_b^{**} = x_g^{**} = x_b^* = x_g^* = 2$, in order to get that $U_b(2) = U_g(2)$ I have that $y_b^0 = y_g^0 = y_{all}^0 = y^0$. This motivates the following theorem:

Theorem 9. *If individuals have complete insurance from health shocks, and the firm earns non-negative profits, then no workers will shirk.*

Proof. Proved above; non-negative profit condition required to ensure that the firm does not shut down. □

2.3.3 Insurance Exchange

The premium in the insurance exchange, R^{EX} , is determined based on the average medical expenditures of all participants in the exchange, multiplied by $1 + \xi$, where $\xi > 0$ is a

loading factor for the health insurance exchange. Participants in the exchange may consist of both employed workers who are not provided health insurance by their employers, and unemployed workers. In other words:

$$R^{EX} = (1 + \xi) \left(\frac{\bar{m}_g n_g^{EX} + \bar{m}_b n_b^{EX}}{n_g^{EX} + n_b^{EX}} \right) \quad (2.3.16)$$

where \bar{m}_h is the average medical expenditure for people of type h health status (the insurers are risk neutral, do not care about the distribution of medical expenditure shocks, whereas individuals do), and n_h^{EX} is the number of participants of type h health status in the exchange.

Notice that choosing to purchase health insurance or remaining unemployed is a *static* decision: it only affects the realization of medical risk for one period after purchase. In other words, an unemployed individual of health status h will want to buy health insurance ($x_h^* = 2$) if and only if:

$$v_h(b, 2) > v_h(b, 0) \iff u(T(b) + S(b, R^{EX}) - R^{EX}) > \mathbb{E}_{\tilde{m}_h}(u(T(b) - \tilde{m}_h - P_W(b))) \quad (2.3.17)$$

Similarly, an employed individual of health status h not offered health insurance by his employer will want to buy health insurance ($x_h^{**} = 2$) if and only if:

$$v_h(y_{eq}, 2) > v_h(y_{eq}, 0) \iff u(T(y_{eq}) + S(y_{eq}, R^{EX}) - R^{EX}) > \mathbb{E}_{\tilde{m}_h}(u(T(y_{eq}) - \tilde{m}_h - P_W(y_{eq}))) \quad (2.3.18)$$

Then, I can develop sufficient conditions for when individuals with bad health elect to purchase health insurance from the market. This result is unique to the world after the

introduction of the ACA:

Theorem 10. $x_g^* = 2 \Rightarrow x_b^* = 2$ and $x_g^{**} = 2 \Rightarrow x_b^{**} = 2$

Proof. Follows from the fact that $v_g(y, 2) = v_b(y, 2) = u(T(y) + S(y, R^{EX}) - R^{EX})$ and $\mathbb{E}_{\tilde{m}_g}(u(T(y) - \tilde{m}_g - P_W(b))) > \mathbb{E}_{\tilde{m}_b}(u(T(y) - \tilde{m}_b - P_W(b)))$. Since this holds for all y , the proof is complete. \square

Intuitively, this result is driven from the fact that discrimination on pre-existing conditions is non-existent in my model. In other words, the cost of obtaining health insurance is the same for individuals of good and bad health, but the benefits are much larger to individuals with bad health. This theorem states that if individuals with good health elect to purchase health insurance, individuals with bad health will too.

I present the following table to illustrate the different possible values of n_h^{EX} when the wage y_{all}^0 is offered in equilibrium:

Table 2.1: Some Possible Values of n_h^{EX}

Case	Equilibrium Wage	Health Status, h	x_h^*	x_h^{**}	n_h^{EX}
1.1	y_{all}^0	b	2	2	B
		g	2	2	G
1.2	y_{all}^0	b	2	2	B
		g	2	0	$\frac{\delta G}{\alpha + \delta}$
1.3	y_{all}^0	b	2	0	$\frac{\delta B}{\alpha + \delta}$
		g	2	0	$\frac{\delta G}{\alpha + \delta}$
1.4	y_{all}^0	b	2	2	B
		g	0	2	$\frac{\alpha G}{\alpha + \delta}$
1.5	y_{all}^0	b	2	2	B
		g	0	0	0
1.6	y_{all}^0	b	2	0	$\frac{\delta B}{\alpha + \delta}$
		g	0	0	0
1.7	y_{all}^0	b	0	2	$\frac{\alpha B}{\alpha + \delta}$
		g	0	2	$\frac{\alpha G}{\alpha + \delta}$
1.8	y_{all}^0	b	0	2	$\frac{\alpha B}{\alpha + \delta}$
		g	0	0	0
1.9	y_{all}^0	b	0	0	0
		g	0	0	0

Moreover, when health insurance is part of the optimal contract offered by the firm, i.e., y^1 is offered, then the value of x_h^{**} becomes meaningless since individuals never demand private health insurance while employed. This cuts down the number of cases when employers offer HI to three, and means I have a total of $3 * 9 + 3 * 3 = 36$ possible cases to consider.

When the equilibrium wage leads to shirking by some individuals, i.e., I have y_g or y_b , the only change will be that n_h , the steady-state level of employment of type h health status workers, will change. In other words, my table would be altered by replacing $\frac{\alpha h}{\alpha + \delta}$

with $\frac{\alpha h}{\alpha + \delta + q}$ and replacing $\frac{\delta h}{\alpha + \delta}$ with $\frac{(\delta + q)h}{\alpha + \delta + q}$, where $h \in \{B, G\}$ and h corresponds to the health-status of workers allowed to shirk.

2.3.4 Solving for No-Shirk-Conditions

The no-shirk conditions after the introduction of the ACA take the same form as the no-shirk conditions prior to the introduction of the ACA. In other words, I get the same no-shirk conditions as in section 2.2.4, the derivation of which is in the appendix.

Case 1: Firm offers HI

Recall the equation from the benchmark model.

$$v(y, 1) = v(b, 1) + \left(\frac{1 - \alpha}{\alpha}\right)v_h(b, x^*) + (1 + \delta + q + r)e/q$$

Now, notice that there is a hidden effect from offering health insurance: doing so will change the composition of individuals in the health insurance exchange, thus changing the price of health insurance in equilibrium, R^{EX} , and affecting x^* .

Again, notice that since only $v(b, x^*)$ depends on health status in the above equality, and $v_g(b, x^*) \geq v_b(b, x^*)$, it is necessarily the case that the NSC of the good-health status worker will be binding: $y_g^1 \geq y_b^1$. This motivates the following theorem:

Theorem 11. *When the firm offers health insurance, the NSC of the good-health status worker will be binding: $y_{all}^1 = y_g^1 \geq y_b^1$.*

Proof. By contradiction, as in Theorem 3. □

I first proceed by solving this simpler problem, where the firm is “forced” to provide health insurance, in order to characterize the workers’ purchase/do not purchase HI while unemployed decision, and relate it to equilibrium wage offer and firm profit. This leads us to the following theorem:

Theorem 12. *When the firm offers HI, there exist parametric regions where $x_g^* = x_b^* = 0$: (**Case 1**), $x_g^* = 0$ and $x_b^* = 2$: (**Case 2**), and $x_g^* = x_b^* = 2$: (**Case 3**). Moreover, there exist parametric regions where both Case 2 and Case 3 can be supported as equilibria. In those cases, the firm will choose the smaller y , which corresponds to (**Case 2**). I refer to this as the “bad” equilibrium, since the good-health workers remain uninsured and do not want to purchase HI since the market is “polluted” with only bad-health individuals.*

Proof. Simulation shows the regions (figures 1 and 2 in the appendix). Letting the subscript denote the case number, I must have that $R_2^{EX} > R_3^{EX}$, since $\bar{m}_b > \bar{m}_g$ and $n_g^{EX} \neq 0$. This implies that $v_2(b, x^*) \leq v_3(b, x^*)$, which drives down the right-hand side of (2.2.26) which, consequently, means that $y_2 \leq y_3$. Since the firm’s only goal is to maximize profit, I have that y_2 and, hence, Case 2, arises in equilibrium when both Case 2 and Case 3 can be supported. \square

Case 2: Firm does not offer HI

Recall the equation from the benchmark model:

$$v(y, x^{**}) = \frac{1}{\alpha} v(b, x^*) + (1 + \delta + q + r)e/q$$

Note that again there is a hidden effect from offering health insurance: doing so will

change the composition of individuals in the health insurance exchange, thus changing the price of health insurance in equilibrium, R^{EX} , and affecting x^* and x^{**} .

Again, notice that since only $v(y, x^{**})$ and $v(b, x^*)$ depend on health status in the above equality, and $v_g(b, x^*) \geq v_b(b, x^*)$, it is necessarily the case that if $x^{**} = 2$, in which case $v_h(y, x^{**}) = v_g(y, x^{**})$, that the NSC of the good-health status worker will be binding: $y_g^1 \geq y_b^1$. This motivates the following theorem, which is the analog of Theorem 11 for the case where the firm does not offer HI:

Theorem 13. *When the firm does not offer health insurance, but the worker self-insures himself while employed ($x^{**} = 2$), the NSC of the good-health status worker will be binding: $y_{all}^0 = y_g^0 \geq y_b^0$.*

Proof. By contradiction, as in Theorem 3. □

Combining Cases

$$v(y^1, 1) - v(y^0, x^{**}) = v(b, 1) - v(b, x^*) \quad (2.3.19)$$

If individuals always choose to self-insure, (2.3.19) becomes:

$$u(y^1) - u(y^0 + S(y^0) - R) = u(b) - u(b + S(b) - R) \quad (2.3.20)$$

Since $y > b \Rightarrow S(y) < S(b)$ (by assumption) $\Rightarrow R - S(y) > R - S(b)$. Since u is concave, it is unclear which of y^1 or y^0 is larger. Notice, however, that if u was linear, I

would get:

$$\begin{aligned}
y^1 - y^0 + (R - S(y^0)) &= R - S(b) \\
\Rightarrow y^1 &< y^0
\end{aligned} \tag{2.3.21}$$

Intuitively, this result tells us that if individuals are risk-neutral, the firm must offer a higher equilibrium wage to get the same no-shirk behavior as when it offered HI, since now the “efficiency wage” has essentially decreased. I can even extend this special case of risk-neutrality to any choice of x^* and x^{**} .

Theorem 14. *If individuals are risk-neutral, then $y^0 > y^1$.*

Proof. In appendix. □

To understand why I am getting an ambiguous result when individuals are risk-averse, consider again the “double-edged sword” nature of offering HI. While offering HI increases utility from not shirking, it *also* increases utility from shirking. This is because individuals get health insurance for a period after being fired.⁷¹ Since this health insurance continuation comes when the individual receives a low income equal to his unemployment insurance benefits, concavity of u implies this health insurance is “worth” more to him (in terms of utility) than before. The reason the effect is unclear is because at this lower income, subsidies to purchase health insurance are also higher, making COBRA continuation less important.

⁷¹As mentioned earlier, I can assume this away by stipulating that if caught shirking, it is a case of “gross misconduct” and, hence, COBRA continuation payments cease.

2.4 Benchmark/ACA Comparison

I begin with a central assumption that motivates the consequent results:

Assumption: $R^{ACA} \geq R_g^B$ and $R^{ACA} \leq R_b^B$, where the superscript B stands for benchmark. What this says is that after the introduction of the ACA, for the same parameter values, good health workers will be charged at least as high a premium as they were paying before (now that they are grouped together in a pool with bad health workers), while bad health workers will be charged no more than they were before (now that they are grouped together in a pool with good health workers).

Then, comparing theorems 1 and 8, I can show the following:

Theorem 15. *If u exhibits CARA,*

$$\begin{aligned} x_b^{*B} = 2 \cup x_b^{**B} = 2 &\Rightarrow x_b^{*ACA} = x_b^{**ACA} = 2 \\ x_b^{*ACA} = 0 &\Rightarrow x_b^{*B} = x_b^{**B} = 0 \cap x_g^{*ACA} = 0 \\ x_b^{**ACA} = 0 &\Rightarrow x_b^{*B} = x_b^{**B} = 0 \cap x_g^{**ACA} = 0 \end{aligned}$$

Proof. In appendix. □

2.4.1 Case 1: Firm offers HI after ACA

If the firm is induced to offer HI after the introduction of the ACA, when previously it had not (this is necessarily the case for CARA or DARA utility), the following must hold

(follows from Equations 2.2.26 and 2.2.27):

$$v^{ACA}(y^{ACA}, 1) - v^B(y^B, x^{B**}) = \frac{1}{\alpha} v^{ACA}(b, x^{ACA*}) - \frac{1}{\alpha} v^B(b, x^{B*}) + v^{ACA}(b, 1) - v^{ACA}(b, x^{ACA*}) \quad (2.4.1)$$

Subcase 1: $x^{*ACA} = 2$, $x^{*B} = x^{**B} = 0$

Theorem 16 (Efficiency Wage Theorem 1). *If u exhibits CARA or DARA, $R - S(b) < b - CE(b, \tilde{m}_h) \forall h$, and the ACA induces the firm to offer HI and individuals to self-insure, whereas this had not previously been the case, then wages will be higher after the implementation of the ACA. (This is somewhat like increasing unemployment benefits in Shapiro and Stiglitz).*

If $R - S(b) < b - CE(b, \tilde{m}_b)$ only, then I require the additional condition that the firm did not offer y_g^B in the benchmark.

If u exhibits CARA and shocks are normally distributed, the above certainty equivalent conditions reduce to $R^{ACA} - S(b) < \bar{m}_h + \frac{1}{2}\gamma \text{Var}(\tilde{m}_h)$.

Proof. Now, if individuals only begin insuring themselves after the implementation of the ACA, and employers offer HI, I must have that:

$$v^{ACA}(y_{h'}^{ACA}, 1) - v_h^B(y_h^B, 0) = \frac{1}{\alpha} v^{ACA}(b, 2) - \frac{1}{\alpha} v_h^B(b, 0) + v^{ACA}(b, 1) - v^{ACA}(b, 2)$$

Notice the distinction between h and h' above. Since $\alpha < 1$ and $v^{ACA}(b, 2) > v_h^B(b, 0)$ (since, by assumption, $b - R + S(b) > CE(b, \tilde{m}_h)$), I must have that $v^{ACA}(b, 2) - v_h^B(b, 0) <$

$\frac{1}{\alpha}[v^{ACA}(b, 2) - v_h^B(b, 0)]$. Then, I get:

$$\begin{aligned}
v^{ACA}(y_{h'}^{ACA}, 1) - v^B(y_h^B, 0) &= \frac{1}{\alpha}v^{ACA}(b, 2) - \frac{1}{\alpha}v^B(b, 0) + v^{ACA}(b, 1) - v^{ACA}(b, 2) \\
&> v^{ACA}(b, 2) - v_h^B(b, 0) + v^{ACA}(b, 1) - v^{ACA}(b, 2) \\
&> v^{ACA}(b, 1) - v_h^B(b, 0) \\
\Rightarrow u(y_{h'}^{ACA}) - \mathbb{E}u(y_h^B - \tilde{m}_h) &> u(b) - \mathbb{E}u(b - \tilde{m}_h) \\
\Rightarrow y_{h'}^{ACA} &> y_h^B
\end{aligned}$$

Where the last line follows from concavity of u and CARA or DARA, and mirrors Theorem 6. (If DARA, $y^B - CE(y^B, \tilde{m}) < b - CE(b, \tilde{m})$ and I can construct a proof by contradiction as before). Notice that the above result holds for any h' , and for the h that the appropriate assumption was made to ensure that $v^{ACA}(b, 2) > v_h^B(b, 0)$. If I assume it for $h = g$, I am done, since $v_g^B(b, 0) > v_b^B(b, 0)$ implies that $v^{ACA}(b, 2) > v_b^B(b, 0)$, which is the condition required to make the proof hold for $h = b$. If I want to make the weaker assumption for $h = b$ only, then I need to make the following more long-winded argument:

So long as the firm sets $y^B = y_b^B$ in the benchmark, I will necessarily have that $y^{ACA} > y^B$ (I am ruling out the case where the firm offers y_g^B in the benchmark since it could, in theory, be larger than the ACA wage offer). A stronger (but more readily observable) condition to ensure this is that good-health workers shirk.

With CARA utility, the condition to ensure that $v^{ACA}(b, 2) > v_h^B(b, 0)$ reduces to:
 $R^{ACA} - S(b) < \bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h)$. □

Subcase 2: $x^{ACA} = 2$, $x^{*B} = 2$

Notice that if $x^{B*} = 2$, all else the same (this can only happen with non-CARA utility, like DARA utility, otherwise I would violate Theorem 1), and $v^{ACA}(b, 2) > v^B(b, 2)$ (which occurs when the effective premium under the ACA when unemployed is lower than the premium in the benchmark: $R^{ACA} - S(b) < R_h$, and is always true (without assumption) for the bad-health worker), then the same proof holds:

$$\begin{aligned}
v^{ACA}(y_{h'}^{ACA}, 1) - v_h^B(y_h^B, 0) &= \frac{1}{\alpha} v^{ACA}(b, 2) - \frac{1}{\alpha} v_h^B(b, 2) + v^{ACA}(b, 1) - v^{ACA}(b, 2) \\
&> v^{ACA}(b, 2) - v_h^B(b, 2) + v^{ACA}(b, 1) - v^{ACA}(b, 2) \\
&> v^{ACA}(b, 1) - v_h^B(b, 2) \\
\Rightarrow u(y_{h'}^{ACA}) - \mathbb{E}u(y^B - \tilde{m}_h) &> u(b) - u(b - R_h) \\
\Rightarrow u(y_{h'}^{ACA}) - u(y_h^B - R_h) &> u(b) - u(b - R_h) \\
\Rightarrow y_{h'}^{ACA} &> y_h^B
\end{aligned}$$

Where the fifth line follows from $x^{B**} = 0 \Rightarrow \mathbb{E}u(y^B - \tilde{m}) > u(y^B - R_h)$ and the last line follows from concavity of u .

Notice, moreover, that if $x^{B**} = 2$ as well (this will necessarily be the case with CARA utility), the proof is identical, with the fourth line omitted. These results motivate the following theorem:

Theorem 17 (Efficiency Wage Theorem 2). *For any concave utility function if, after the implementation of the ACA, firms offer HI, individuals always self-insure (and in they benchmark they insured while unemployed), and effective premiums when unemployed are*

lower for the good-health individual ($R^{ACA} - S(b) < R_g$), then wages will be higher after the implementation of the ACA.

If effective premiums when unemployed are not lower for the good-health individual, then wages will still be higher after the implementation of the ACA provided that the firm did not offer y_g^B in the benchmark (again, a stronger condition that ensures this is that good-health workers shirked in the benchmark).

Proof. Same as before. □

2.4.2 Case 2: Firm does not offer HI

Next, similar to what I did in the benchmark case, I can derive, explicitly, the wage that will cause the NSC to bind conditional on the health insurance decision of the workers, and the health insurance offer decision of the firm, after the implementation of the ACA. Then, I can compare the wages before and after the implementation of the ACA to determine how efficiency wages will change as a result of the ACA.

The following theorems are proved exclusively for utility that exhibits CARA.

Recall the NSC when the firm does not offer health insurance: $v(y, x^{**}) = \frac{1}{\alpha}v(b, x^*) + (1 + \delta + q + r)e/q$. Let us use this to solve for y_g^0 and y_b^0 , as before. I will henceforth omit the superscript *ACA* and refer to the wages under the ACA as y_h^0 or y_h^1 , and refer to the wages under the benchmark as y_h^B (since the firm will never offer health insurance in the benchmark).

Subcase 1: $x_h^{*ACA} = x_h^{**ACA} = 2$

Then, letting $(1 + \delta + q + r)e/q = Q$, as before, the NSC reduces to:

$$\begin{aligned}
v(y, 2) &= \frac{1}{\alpha}v(b, 2) + (1 + \delta + q + r)e/q \\
u(T(y) - R + S(y)) &= \frac{1}{\alpha}u(T(b) - R + S(b)) + Q \\
-\alpha e^{-\gamma[T(y) - R + S(y)]} &= -e^{-\gamma[T(b) - R + S(b)]} + \alpha Q \\
\alpha e^{-\gamma[T(y) + S(y)]} &= e^{-\gamma[T(b) + S(b)]} - \alpha Q e^{-\gamma R} \\
\log(\alpha) - \gamma[T(y) + S(y)] &= \log(e^{-\gamma[T(b) + S(b)]} - \alpha Q e^{-\gamma R}) \\
T(y_h^0) &= \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma[T(b) + S(b)]} - \alpha Q e^{-\gamma R}}\right) - S(y_h^0) \\
&= \frac{1}{\gamma} \log\left(e^{\gamma S(b)} \frac{\alpha}{e^{-\gamma T(b)} - \alpha Q e^{-\gamma R} e^{\gamma S(b)}}\right) - S(y_h^0) \\
&= \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma T(b)} - \alpha Q e^{-\gamma[R - S(b)]}}\right) + [S(b) - S(y_h^0)]
\end{aligned}$$

Recall that in the benchmark model, when individuals chose to self-insure at income b and y (i.e., when $x_h^* = x_h^{**} = 2$), I had that:

$$T(y_h^B) = \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma T(b)} - \alpha Q e^{-\gamma R_h}}\right)$$

Since $y_h^0 > b$, I have that $S(b) - S(y_h^0) > 0$. This motivates the following theorem:

Theorem 18. *Given CARA utility, if the firm does not offer HI, individuals always self-insure, and $R^{ACA} - S(b) \leq R_g^B$, then $y^{ACA} > y^B$. If $R^{ACA} - S(b) > R_g^B$, but the firm does not set y_g^B in the benchmark, then $y^{ACA} > y^B$.*

Proof. Theorems 4 and 9 tell us that, in this situation, the NSC of the good health worker

will be binding in the benchmark, i.e., $y_g^B > y_b^B$ and the NSC's will be identical under the ACA, i.e., $y_g^0 = y_b^0$. Since $R^{ACA} \leq R_b^B$ by assumption, I have that $R^{ACA} - S(b) \leq R_b^B$, so $T(y_h^0) > T(y_b^B)$ for $h \in \{g, b\}$. This implies that $y_b^0 > y_b^B$ and $y_g^0 > y_b^B$. So if the firm did not set y_g^B in the benchmark (or, alternatively, if the firm allowed the good-health workers to shirk in the benchmark), the equilibrium wage after the implementation of the ACA, given by either y_b^0 or y_g^0 , will be strictly greater than what it was before the implementation of the ACA, y_b^B .

If, on the other hand, I also have that the effective premium paid by the good-health workers is lower under the ACA, or $R^{ACA} - S(b) \leq R_g^B$, then I will also have that $y_g^0 > y_g^B$ and $y_b^0 > y_g^B$, so no matter which NSC's hold, the equilibrium wage will be higher than in the world without the ACA. \square

If individuals always self-insure, this theorem tells us that if, in the world with the ACA, the effective premium that individuals of health status g have to pay when unemployed (premium minus subsidies) is no larger than what they had to pay before the ACA, the equilibrium wage will increase.

If the effective premium is larger, but the firm only ensured the NSC of the bad-health workers held in the benchmark (i.e., it allowed the good health workers to shirk), then the equilibrium wage will also increase under the ACA.

Corollary 3. *If $R^{ACA} - S(b) - P(b) \leq R_g^B$, then $(x_g^{*B} = 2 \cup x_g^{**B} = 2) \Rightarrow x_g^{*ACA} = 2$*

*If $R^{ACA} - S(y) - P(y) \leq R_g^B$, then $(x_g^{*B} = 2 \cup x_g^{**B} = 2) \Rightarrow x_g^{**ACA} = 2$*

*If $R^{ACA} - S(y) \leq R_g^B$, then $(x_g^{*B} = 2 \cup x_g^{**B} = 2) \Rightarrow x_g^{*ACA} = x_g^{**ACA} = 2$*

Proof. For a proof by contradiction, consider $R^{ACA} - S(b) - P(b) \leq R_g^B$ and $x_g^{*B} = 2$ but

$x_g^{*ACA} = 0$. Then, from Theorem 1, I have that:

$$x_g^{*B} = 2 \iff x_g^{**B} = 2 \iff R_g^B < \frac{1}{\gamma} \log \int e^{\gamma x} f_g(x) dx$$

And from Theorem 8, I have that:

$$\begin{aligned} x_g^{*ACA} = 0 & \iff R^{ACA} > \frac{1}{\gamma} \log \int e^{\gamma x} f_g(x) dx + S(b) + P(b) \\ & \Rightarrow R_g^B \geq R^{ACA} - S(b) - P(b) > \frac{1}{\gamma} \log \int e^{\gamma x} f_g(x) dx \\ & \Rightarrow R_g^B > \frac{1}{\gamma} \log \int e^{\gamma x} f_g(x) dx \\ & \Rightarrow x_g^{*B} = 0 \end{aligned}$$

Which yields the desired contradiction. Repeat for income of y to get the second result.

The final result follows since $R^{ACA} - S(b) - P(b) \leq R^{ACA} - S(y)$ and $R^{ACA} - S(y) - P(y) \leq R^{ACA} - S(y)$. \square

Putting all these theorems together, I can present the third Efficiency Wage Theorem:

Theorem 19 (Efficiency Wage Theorem 3). *If the firm does not offer HI, individuals self-insured before the introduction of the ACA ($\xi_h < \frac{1}{\gamma \bar{m}_h} \log \int e^{\gamma x} f_h(x) dx - 1$), and the effective premium is lower under the ACA for the good-health worker when employed ($R^{ACA} - S(y) \leq R_g^B$),⁷² then wages are higher after the implementation of the ACA.*

If the effective premium is higher under the ACA for the good-health workers, but they still always self-insure, and the NSC of the good-health workers does not hold ($y < y_g$),

⁷²I actually only require the weaker (but less intuitive) conditions that $R^{ACA} - S(y) - P(y) \leq R_g^B$ and $R^{ACA} - S(b) \leq R_g^B$.

at least in the benchmark, then wages will still be higher after the implementation of the ACA.

Proof. Since individuals self-insured before the introduction of the ACA (i.e., $x_h^{B*} = x_h^{**B} = 2$), Theorem 15 tells us that $x_b^{*ACA} = x_b^{**ACA} = 2$. Moreover, since the effective premium is lower under the ACA for the good health worker when employed, it follows from Corollary 3 that $x_g^{*ACA} = x_g^{**ACA} = 2$. In other words, it must be the case that individuals completely self-insure even after the implementation of the ACA. The first result then follows directly from Theorem 18.

Next, notice that the bad-health workers will always self-insure after the implementation of the ACA if they did so before its implementation (Theorem 15). So, as long as the firm does not set y_g^B in the benchmark, we get that $y^{ACA} > y^B$. This second result follows directly from Theorem 18. \square

Subcase 2: $x_h^{*ACA} = 0$ and $x_h^{**ACA} = 2$

Again, letting $(1 + \delta + q + r)e/q = Q$, the NSC reduces to:

$$\begin{aligned} v(y, 2) &= \frac{1}{\alpha}v(b, 0) + (1 + \delta + q + r)e/q \\ u(T(y) - R + S(y)) &= \frac{1}{\alpha}u(T(b) - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h) - P(b)) + Q \\ T(y_h^0) &= \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma[T(b) - P(b) - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h)]} - \alpha Q}\right) + R - S(y_h^0) \\ T(y_h^0) &= \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma[T(b) + (R - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h) - P(b) - S(y))]} - \alpha Q e^{-\gamma[R - S(y)]}}\right) \end{aligned}$$

Recall that in the benchmark model, when individuals chose to self-insure at income

b and y (i.e., when $x_h^* = x_h^{**} = 2$), I had that:

$$T(y_h^B) = \frac{1}{\gamma} \log\left(\frac{\alpha}{e^{-\gamma T(b)} - \alpha Q e^{-\gamma R_h}}\right)$$

Now, if $R - S(y) < R_g$, I get that $y_g^0 > y_g^B$. To see this, notice that since $x_h^{*ACA} = 0$, I have that $R^{ACA} > S(b) + P(b) + \bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h) > S(y) + P(b) + \bar{m}_h + \frac{1}{2}\gamma Var(\tilde{m}_h) \Rightarrow (R - \bar{m}_h - \frac{1}{2}\gamma Var(\tilde{m}_h) - P(b) - S(y)) > 0$. Notice, however, that I cannot do the same trick for y_b^0 since, by Theorem 15, $x_b^{*ACA} = 0 \Rightarrow x_b^{*B} = 0$, which makes the comparison of the above wage equations inappropriate. It is, though, entirely plausible to have $x_g^{*ACA} = 0$, $x_g^{**ACA} = 2$, $x_g^{*B} = 2$, and $x_g^{**B} = 2$.

Theorem 20 (Efficiency Wage Theorem 4). *If the firm does not offer HI, and good-health individuals stop insuring themselves when unemployed after the introduction of the ACA but continue insuring themselves while employed, at a lower effective premium, then $y_g^{ACA} > y_g^B$. If, moreover, good-health individuals do not shirk, then wages are higher after the implementation of the ACA.*

If the above conditions hold, but good-health individuals shirk, then wages are still higher after the implementation of the ACA provided that bad-health individuals only insure themselves when employed after the introduction of the ACA.

Proof. The conditions of the first statement are equivalent to $x_g^{*ACA} = 0$, $x_g^{**ACA} = 2$, $x_g^{*B} = 2$, $x_g^{**B} = 2$ and $R^{ACA} - S(y) < R_g$ which, as I already showed above, lead to $y_g^{ACA} > y_g^B$. By Theorem 4, I know that $y_g^B > y_b^B$. So, as long as the firm ensures that good-health workers do not shirk (i.e., it gives them a wage greater than or equal to y_g^{ACA} , that wage will necessarily be larger than any possible wage in the benchmark. In other

words, this gives the result that wages rise after the implementation of the ACA.

On the other hand, if good-health workers shirk but bad-health individuals only insure themselves when employed after the introduction of the ACA, I have $x_b^{*ACA} = 0$ and $x_b^{**ACA} = 2$. Then I can use the above equation, as I did before, to show that $y_b^{ACA} > y_g^B$. Since $y_g^B > y_b^B$, I have that wages will increase after the introduction of the ACA. \square

Putting all these theorems together, I have that for nearly all cases of interest (when the ACA succeeds in doing what it is supposed to do: induce firms to offer HI or unhealthy individuals to self-insure, or any combination thereof) wages will increase so long as effective premiums are lower under the ACA.

In other words, if subsidies under the ACA are such that the effective premium is lowered, the ACA will (in most cases of interest), cause efficiency wages to increase. What does this mean for consumer welfare?

Welfare Analysis

As in Shapiro and Stiglitz [1984], I begin with the case where the owners of the firms are the same individuals as the workers, and ownership is equally distributed among the $B + G$ workers. Notice that none of the results in this section depended on the production function of the firm. Indeed, to follow Shapiro and Stiglitz's result, I will modify the production function to be a concave and increasing function of the labor force: $F(n_g, n_b)$, such that full employment is optimal $F'(G, B) > e$ (my constant returns to scale function would not generate the result below, see Shapiro and Stiglitz [1984] for more details). Furthermore, notice that since the firm is risk-neutral, it is socially optimal for it to offer HI, provided the utility function of the agents is not risk loving. Then, I can write the

planners problem that maximizes the total utility of all workers as:

$$\begin{aligned}
\max_{y, b, n_g, n_b, S(b), P(b)} \quad & (u(y) - e)(n_g + n_b) + v_b(b, x_b^*)(B - n_b) + v_g(b, x_g^*)(G - n_g) \quad (2.4.2) \\
\text{s.t} \quad & v(y, 1) \geq v(b, 1) + \left(\frac{1 - \alpha}{\alpha}\right)v_h(b, x_h^*) + (1 + \delta + q + r)e/q \text{ for } h = b, g \quad (\text{NSC's}) \\
\text{s.t} \quad & y[n_g + n_b] + b[B + G - n_g - n_b] + n_g(\bar{m}_g) + n_b(\bar{m}_b) \leq F(n_g, n_b) \quad (\text{Firm Feasibility}) \\
\text{s.t} \quad & S(b) \leq P(b) \quad (\text{Gov. Feasibility})
\end{aligned}$$

Notice that the choice of n_g and n_b affect the health insurance exchange which affects the worker's self-insurance choice, x_h^* , which affects his instantaneous utility, $v_h(b, x_h^*)$.

Furthermore, notice that α , the rate of obtaining a job, can be related to more fundamental parameters of the model. In particular, since the flow in must equal the flow out, I have that, when both NSC's are satisfied:

$$\begin{aligned}
\alpha(B + G - n_b - n_g) &= \delta(n_g + n_b) \\
\Rightarrow \alpha &= \frac{\delta(n_g + n_b)}{B + G - n_b - n_g} \\
\alpha &= \frac{\delta L}{N - L}
\end{aligned}$$

Where, following Shapiro and Stiglitz [1984], I have let $B + G = N$ and $n_g + n_b = L$. Notice that $1/\alpha$ is the expected duration of being unemployed, and that *no shirking is inconsistent with full employment*; if $L = N$, then everyone would shirk, knowing they would immediately get re-hired if they lost their jobs (since $\alpha = +\infty$). Now, my NSC's are functions of the employment rate (FOC's are in the appendix, for future work).

I can write the competitive problem for the firm that offers health insurance and

ensures both workers do not shirk as:

$$\begin{aligned} \max_{y, b, n_g, n_b} \quad & F(n_g, n_b) - y[n_g + n_b] + b[B + G - n_g - n_b] + n_g(\bar{m}_g) + n_b(\bar{m}_b) \\ \text{s.t.} \quad & v(y, 1) \geq v(b, 1) + \left(\frac{1-\alpha}{\alpha}\right)v_h(b, x_h^*) + (1 + \delta + q + r)e/q \text{ for } h = b, g \quad (\text{NSC's}) \end{aligned}$$

Since this problem ignores the welfare maximization of the workers (whose objective function is strictly increasing in the wage level), it will necessarily have a wage no larger than what is socially optimal. I can similarly write the competitive problem for the firm that does not offer health insurance. However, I need not write the problem where only one group of workers' NSC conditions hold, since the wage that satisfies this will necessarily be smaller or equal to the wage that ensures both NSC conditions hold. In other words, I have (informally) shown that when the firm offers health insurance, it will offer a wage that is less than or equal to the socially optimal wage level, no matter which NSC's hold. It will take more work to show this for the case when the firm does not offer health insurance.

A simpler (intuitive, but not completely correct) argument is that the firm will choose to hire L workers and offer them the wage y such that $F'(L) = y$ and $u(y) =$ LHS of binding NSC. Thus, the market outcome will be at a wage, y , that equals the marginal productivity of labor, while the social optimum will be at a wage, y^* , that equals the average product of labor. Since $F(L)$ is concave, I have that $y^* > y$, and the market outcome is sub-optimal. Indeed, government intervention that raises the equilibrium wage leads us in the right direction and may provide Pareto improvements.

Theorem 21. *The market outcome is socially inefficient. The optimal wage, y^* , is greater*

than the market equilibrium wage, y (when the firm offers health insurance). The ACA raises wages in many parametric regions, which can be seen as efficiency wage subsidies towards the socially optimal wage level.

Proof. Informal discussion above, and from Efficiency Wage Theorems 1-4. □

2.5 Conclusion

In this chapter, I extended the basic model of efficiency wages to incorporate employer-sponsored health insurance. Analogous to the finding in Shapiro and Stiglitz [1984], I show that the market equilibrium is not efficient, and that there are circumstances in which the government's introduction of the Affordable Care Act is welfare-enhancing. There is a parallel between this result and the result in Shapiro and Stiglitz [1984]: the ACA reduces the cost of unemployment by lowering premiums in the health insurance market, much like unemployment insurance does. There are also some differences: the mechanism whereby the ACA leads to efficiency wage subsidies is more complex, and this is primarily driven by heterogeneity in health status. In particular, as subsidies increase, employers are more inclined to offer HI to their workers. This changes the composition of health status of individuals using the insurance exchanges, driving premiums down. In turn, the relative cost of unemployment goes down, forcing firms to increase wages in order to induce good (no-shirk) behavior.

In effect, I have shown how the ACA reform will affect market wages in a simple economy with imperfect monitoring, cost-of-effort, and workers heterogeneity in health status. As an extension, one could consider the problem where workers are heterogeneous

above and beyond their health status. The problem then is that if workers are not all identical, being fired will carry a stigma (new employers can infer something about the fired worker's health status), and firms may adjust their wage offers accordingly. In other words, in such a model, worker heterogeneity would introduce *adverse selection*, above and beyond the *moral hazard* present in this chapter. One could also introduce firm heterogeneity.

Ultimately, the theory developed in this chapter has a meaningful application: a quantitative analysis of the ACA. With some simplifications, this model could be combined with data on employment and health insurance (e.g. the CPS, ASEC Supplement, and MEPS) to deliver a welfare analysis and counterfactual predictions pertaining to the introduction of the ACA. I leave this for future work.

3 Mirrlees Taxation with Endogenous Search Effort

3.1 Introduction

This is the second paper motivated by the hump-shaped lifecycle profile of job search identified by Aguiar et al. [2013]. This chapter extends Mirrlees' theory of optimal taxation (Mirrlees [1971]) to a framework with endogenous job search, and applies it to the estimation of optimal federal tax rates in Canada.

There is a large literature on optimal tax rates, beginning with Mirrlees [1971]. Rather than perform a detailed analysis of this extensive literature, I point the reader to a summary by Mankiw et al. [2009], and highlight only a few particularly relevant papers. Consistent with Mirrlees [1971], Sadka [1976] and Seade [1977] found that the marginal tax rate should be zero at the income level of the top income level, when the income distribution is bounded. This chapter is consistent with this result: for the highest tax bracket, the optimal federal tax rates are indeed zero. More recently, Saez [2001] has argued that elasticities are the key in calculating optimal income tax rates, and determined that the optimal tax rate for labor income needs to be between 50 and 80 percent. This chapter finds that to be the optimal tax rate for the second highest tax bracket.

There is also a large literature on endogenous job search, pioneered by Burdett,

Mortenson, and Pissarides. For a nice overview, see, in order: Burdett [1978], Mortensen [1986] and Mortensen and Pissarides [1994]. In this chapter, workers' search intensity is endogenous as in Christensen et al. [2005]. While this chapter does not study sorting (in this chapter, workers are homogenous), there is a growing literature on endogenous search and sorting; see Lentz and Mortensen [2010] and Bagger and Lentz [2014].

The chapter is organized as follows. Section 2 outlines the economic model. Section 3 details the data. Section 4 discusses the estimation approach, identification issues, and results. Section 5 concludes.

3.2 Model

Consider the prototypical, partial equilibrium search model of an individual with endogenous search effort. Let λs be the flow rate of job offers when unemployed and $\lambda s(w)$ be the flow rate of job offers when employed at wage w , where s and $s(w)$ are the endogenous levels of search effort when unemployed and employed at wage w , respectively. Individuals all share the same cost function $c(s)$. Let $c(s)$ be concave and increasing with $c(0) = 0$ and $\lim_{c \rightarrow \infty} c'(s) = 0$. For now, the cost of searching is the same while employed and unemployed, but I can later generalize the model to make them different.⁷³

The government imposes a net tax on individuals according to the function $T(w)$ when employed at wage w , and T_u when unemployed. Note that either net tax could, in fact, be a subsidy. The government chooses the optimal tax rate in order to maximize a social welfare function, which will have some curvature (inequality aversion) over all possible states (employed at all wages w and unemployed), weighted by the steady state fraction

⁷³One might think that it is easier for employed workers to find a job, as they have an "in", but one may also think it is harder, since they have less free time.

of individuals in those respective states.

3.2.1 Worker's Problem

The value function of unemployment can be written as:

$$\begin{aligned}\rho V_u &= \max_s \{b - T_u + \lambda s \int (\max\{V_e(x), V_u\} - V_u) dF(x) - c(s)\} \\ &= b - T_u + \max_s \left\{ \lambda s \int_{w_R} (V_e(x) - V_u) dF(x) - c(s) \right\}\end{aligned}$$

The value function of employment at the wage w can be written as:

$$\begin{aligned}\rho V_e(w) &= \max_{s(w)} \{w - T(w) + \lambda s(w) \int (\max\{V_e(x), V_e(w)\} - V_e(w)) dF(x) - c(s(w)) + \delta(V_u - V_e(w))\} \\ &= w - T(w) + \max_s \left\{ \lambda s(w) \int_w (V_e(x) - V_e(w)) dF(x) - c(s(w)) \right\} + \delta(V_u - V_e(w))\end{aligned}$$

Notice that it need not be the case that $w_R = b + (T_e(w_R) - T_u)$, since search is costly, though we will show that this is the case.

Let the optimal search effort be s^* when unemployed and $s^*(w)$ when employed. Then, I can derive the reservation wage as follows:

$$\begin{aligned}\rho V_u &= b - T_u + \lambda s^* \int_{w_R} (V_e(x) - V_u) dF(x) - c(s^*) \\ &= b - T_u + \lambda s^* [(V_e(x) - V_u) \bar{F}(w)]|_{w_R}^{\bar{w}} + \lambda s^* \int_{w_R} \bar{F}(x) dV_e(x) - c(s^*) \\ &= b - T_u + \lambda s^* \int_{w_R} \bar{F}(x) dV_e(x) - c(s^*)\end{aligned}$$

Where the second line follows from integration from parts after noticing that $dF(w) = d(1 - F(w)) = d\bar{F}(w)$. Similarly:

$$\begin{aligned}\rho V_e(w) &= w - T(w) + \lambda s^*(w) \int_w (V_e(x) - V_e(w)) dF(x) - c(s^*(w)) + \delta(V_u - V_e(w)) \\ &= w - T(w) + \lambda s^*(w) \int_w \bar{F}(x) dV_e(x) - c(s^*(w)) + \delta(V_u - V_e(w))\end{aligned}$$

Taking derivatives of this expression with respect to w , I find (key: use the envelope theorem, do not take into account derivative of s or $c(s)$ with respect to w since when we maximize, we set this derivative equal to zero):

$$\begin{aligned}\rho V'_e(w) &= 1 - T'(w) - \lambda s^*(w) \bar{F}(w) V'_e(w) - \delta V'_e(w) \\ \Rightarrow V'_e(w) &= \frac{1 - T'(w)}{\rho + \delta + \lambda s^*(w) \bar{F}(w)}\end{aligned}$$

Plugging this expression into the value functions of unemployment and employment, respectively, gives:

$$\rho V_u = b + \lambda s^* \int_{w_R} \frac{(1 - T'(x)) \bar{F}(x) dx}{\rho + \delta + \lambda s^*(x) \bar{F}(x)} - c(s^*) \quad (3.2.1)$$

$$\rho V_e(w) = w + \lambda s^*(w) \int_w \frac{(1 - T'(x)) \bar{F}(x) dx}{\rho + \delta + \lambda s^*(x) \bar{F}(x)} - c(s^*(w)) + \delta(V_u - V_e(w)) \quad (3.2.2)$$

Since, by definition:

$$\begin{aligned}
\rho V_u &= \rho V_e(w_R) \\
&= w_R - T(w_R) + \lambda s^*(w_R) \int_{w_R} \frac{(1 - T'(x))\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} - c(s^*(w_R)) + \delta(V_u - V_e(w_R)) \\
&= w_R - T(w_R) + \lambda s^*(w_R) \int_{w_R} \frac{(1 - T'(x))\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} - c(s^*(w_R)) \\
w_R &= b + (T(w_R) - T_u) + \lambda(s^* - s^*(w_R)) \int_{w_R} \frac{(1 - T'(x))\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} + c(s^*(w_R)) - c(s^*)
\end{aligned}$$

From (3.2.1) and (3.2.2), it is easy to see that the optimal search effort much satisfy:

$$c'(s^*) = \lambda \int_{w_R} \frac{(1 - T'(x))\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} \quad (3.2.3)$$

$$c'(s^*(w)) = \lambda \int_w \frac{(1 - T'(x))\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} \quad (3.2.4)$$

Since the integrand of the RHS of (3.2.4) is always positive, the RHS is increasing in w . Since c is assumed to be concave, this implies that the optimal search effort is decreasing in w . This is intuitive; as a higher wage job is secured, the individual has less incentive to search for a better job, since the proportion of higher-wage jobs in the economy decreases (note that utility is linear here, so concavity of the utility function does not play a role).

Notice that under this simple specification, $s^* = s^*(w_R)$ and $c(s^*(w_R)) = c(s^*)$, so $w_R = b$. Later, I can change the exogenous component of the arrival rate when unemployed to $\kappa\lambda$, in order to get a difference between s^* and $s^*(w_R)$.

Steady-State

Equating flow in with flow out of *unemployment*, I get the following:

$$\begin{aligned}\lambda s^* u \bar{F}(w_R) &= \delta(1 - u) \\ \Rightarrow u &= \frac{\delta}{\delta + \lambda s^* \bar{F}(w_R)}\end{aligned}$$

Let $(1-u)G(w)$ be the number of individuals employed at a wage less than w . Equating flow in with flow out of *employment at wage less than w* , I get:

$$\begin{aligned}\lambda u s^* [F(w) - F(w_R)] &= \delta(1 - u)G(w) + \lambda(1 - u)\bar{F}(w) \int^w s^*(x)g(x)dx \\ \Rightarrow g(w) &= \frac{\lambda u s^* [F(w) - F(w_R)] - \delta(1 - u) \int^{w-\epsilon} g(x)dx - \lambda(1 - u)\bar{F}(w) \int^{w-\epsilon} s^*(x)g(x)dx}{\delta(1 - u) + \lambda(1 - u)\bar{F}(w)s^*(w)}\end{aligned}$$

3.2.2 Government's Problem

The government's problem is as follows:

$$\begin{aligned}\max_T \quad & uV_u + (1 - u) \int V_e(x)g(x)dx \\ \text{s.t.} \quad & R(T) \geq c\end{aligned}$$

Where government revenue, $R(T)$ is given by:

$$R(T) = uT_u + (1 - u) \int g(x)T(x)$$

In other words, the government levies a net tax on employed workers $T(w)$ and a tax on unemployed workers T_u such that welfare is maximized subject to a revenue constraint.

3.3 Data

The data used in this chapter comes from the Canadian Labour Force Survey. It is a public-use microdata file that provides a snapshot of the state of the Canadian labor force in May 2014. While it contains over one hundred thousand individuals, the dataset is *not* a panel data, and, as such, Limited Information Maximum Likelihood Estimation (MLE) must be used in estimation. The data set observes workers within an interval $[t_0, t_1]$, and provides:

1. Elapsed unemployment duration at t_0 : τ_0
2. Residual unemployment duration after t_0 : τ_1 (note that $\tau_1 \leq t_1 - t_0$)
3. Accepted wage w at $t_0 + \tau_1$ if individual leaves unemployment by end of recording period

The dataset additionally contains some retrograde information, such as length of *previous* unemployment spell, and whether there was job destruction or voluntary quits. I do not use this information in estimation. Below is a summary of the data used.

Table 3.1: Data Summary by Province

	# UE	# E	Wage (\$1000s)	Time UE	Time E
Ontario	1882	15218	47	6.6	90.8
Quebec	1373	8977	42	6.2	92.2
Alberta	727	6052	57	6.2	68.2
British Columbia	812	5909	46	6.1	81.3
Manitoba	636	5011	43	6.3	83.7
Saskatchewan	450	3770	50	6.7	83.8
Nova Scotia	466	2606	42	6.3	94.1
New Brunswick	478	2511	41	6.3	97.9
Newfoundland	447	1768	50	5.9	87.2
Prince Edward Island	301	1409	40	6.9	85

Provincial tax rates exhibit a lot of heterogeneity. Federal and Provincial tax rates for May 2014 were obtained from the Canada Revenue Agency. I summarize the *marginal* tax rates of the three most populated provinces in the table below. Not shown is the marginal tax rate in Alberta — a flat 10%.

Table 3.2: 2014 Tax Rate Schedule - Canada

	(\$)	Quebec (%)	(\$)	Ontario (%)	(\$)	BC (%)	(\$)	Federal (%)
Tier 1	41495	16	40120	5	37606	5	43953	15
Tier 2	41490	20	40122	9	37607	8	43954	22
Tier 3	17985	24	69758	11	11141	11	48363	26
Tier 4	-	26	70000	12	18504	12	-	29
Tier 5	-	-	-	13	45142	15	-	-
Tier 6	-	-	-	-	-	17	-	-

I plot the provincial tax functions for the four most populated provinces below.

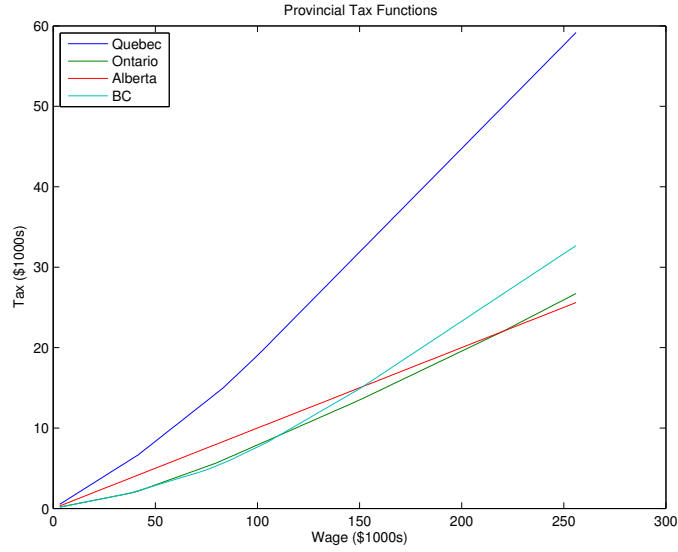


Figure 3.1: Provincial Tax Functions

3.4 Estimation

3.4.1 Limited Information MLE

Given the set-up above, durations are distributed according to the exponential distribution:

$$G(t) = 1 - \exp(-\lambda s^* t)$$

$$g(t) = \lambda s^* \exp(-\lambda s^* t)$$

For workers **unemployed** at time of first interview:

$$\frac{\delta}{\delta + \lambda \bar{F}(w_R) s^*} \times (\lambda \bar{F}(w_R) s^*)^{2-d_{0b}-d_{0f}} \exp(-\lambda \bar{F}(w_R) s^* (t_{0b} + t_{0f})) \times \frac{f(w_0)}{\bar{F}(w_R)}^{1-d_{0f}}$$

Let me explain each component:

1. $\frac{\delta}{\delta + \lambda \bar{F}(w_R)s^*}$ is the probability of being unemployed (steady-state unemployment rate).
2. $(\lambda \bar{F}(w_R)s^*) \exp(-\lambda \bar{F}(w_R)s^* t_{0b})$ is the probability of an unemployment spell lasting t_{0b} . If there is left-censoring (i.e., we do not observe the full length t_{0b} , then all we know is that the unemployment spell is greater than t_{0b} in length, so we use the CDF: $\exp(-\lambda \bar{F}(w_R)s^* t_{0b})$).
3. $(\lambda \bar{F}(w_R)s^*) \exp(-\lambda \bar{F}(w_R)s^* t_{0f})$ is the probability of an unemployment spell lasting t_{0f} . Same logic for right-censoring.
4. $\frac{f(w_0)}{\bar{F}(w_R)}$ is the probability of accepting the wage w_0 (i.e., conditional on getting an acceptable wage, $w > w_R$, the probability of getting the wage w_0).

For workers **employed** at time of first interview:

$$\frac{\lambda \bar{F}(w_R)s^*}{\delta + \lambda \bar{F}(w_R)s^*} \times g(w_1) \times (\delta + \lambda s^*(w_1) \bar{F}(w_1))^{1-d_{1b}} \dots$$

$$\exp[-(\delta + \lambda s^*(w_1) \bar{F}(w_1))(t_{1b} + t_{1f})] \times [\delta^{v_1} (\lambda s^*(w_1) \bar{F}(w_1))^{1-v_1}]^{1-d_{1f}}$$

Let me explain each component:

1. $\frac{\lambda \bar{F}(w_R)s^*}{\delta + \lambda \bar{F}(w_R)s^*}$ is the probability of being employed (steady-state employment rate).
2. $g(w_1)$ is (conditional on being employed) the probability of being employed at a wage w_1 .
3. $(\delta + \lambda s^*(w_1) \bar{F}(w_1)) \exp[-(\delta + \lambda s^*(w_1) \bar{F}(w_1)) t_{1b}]$ is the probability of an employment spell lasting t_{1b} . If there is left-censoring (i.e., we do not observe the full length t_{1b}),

then all we know is that employment spell is greater than t_{1b} in length, so we use the CDF: $\exp[-(\delta + \lambda s^*(w_1)\bar{F}(w_1))t_{1b}]$.

4. $(\delta + \lambda s^*(w_1)\bar{F}(w_1)) \exp[-(\delta + \lambda s^*(w_1)\bar{F}(w_1))t_{1f}]$ is the probability of an employment spell lasting t_{1f} . Same logic for right-censoring.
5. $\frac{\delta}{\delta + \lambda s^*(w_1)\bar{F}(w_1)}$ is (conditional on observing a transition) the probability of job destruction.
6. $\frac{\lambda s^*(w_1)\bar{F}(w_1)}{\delta + \lambda s^*(w_1)\bar{F}(w_1)}$ is (conditional on observing a transition) the probability of a job-to-job transition ("J2J").

Why is it limited-information?

1. Only using info up to first transition
2. No info used on wages following J2J transition (in data could have $w_2 < w_1$, leading to degenerate likelihood)

With the dataset at hand, I additionally do not see the right side of the employment or unemployment spells (since I only see a snapshot of individuals - there is no panel data element). This causes the likelihood to reduce to the following:

For workers **unemployed** at time of first interview:

$$\frac{\delta}{\delta + \lambda \bar{F}(w_R)s^*} \times (\lambda \bar{F}(w_R)s^*)^{1-d_{0b}} \exp(-\lambda \bar{F}(w_R)s^*t_{0b})$$

And for workers **employed** at time of first interview:

$$\frac{\lambda \bar{F}(w_R) s^*}{\delta + \lambda \bar{F}(w_R) s^*} \times g(w_1) \times (\delta + \lambda s^*(w_1) \bar{F}(w_1))^{1-d_{1b}} \exp[-(\delta + \lambda s^*(w_1) \bar{F}(w_1)) t_{1b}]$$

As in Christensen et al. [2005], I assume the cost function to be of the following parametric form:

$$c(s) = c_0 \frac{s^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$$

Using the above cost function, the search intensity equation (3.2.4) reduces to:

$$\begin{aligned} c'(s^*(w)) &= \lambda \int_w \frac{\bar{F}(x) dx}{\rho + \delta + \lambda s^*(x) \bar{F}(x)} \\ c_0 s^*(w)^{\frac{1}{\gamma}} &= \lambda \int_w \frac{\bar{F}(x) dx}{\rho + \delta + \lambda s^*(x) \bar{F}(x)} \\ s^*(w) &= \left(\frac{\lambda}{c_0} \int_w \frac{\bar{F}(x) dx}{\rho + \delta + \lambda s^*(x) \bar{F}(x)} \right)^\gamma \end{aligned}$$

The above is a Volterra equation of the second kind. To solve it numerically, I solve it backwards using a grid of wage values, noting that at the highest wage offered \bar{w} , the integral “disappears”.

3.4.2 Identification

This identification discussion closely follows Christensen et al. [2005]. Most parameters of interest can be identified via maximum likelihood. There are some, however, that cannot.

For example, λ and c_0 are not separately identified. To see this, notice that λ and s^* always appear together in the likelihood function, and that $s^*(w) = (\frac{\lambda}{c_0} \int_w \frac{\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x) \bar{F}(x)})^\gamma \propto h(\lambda/c_0)$. In particular, what is identified is $\lambda s^*(w) \propto \lambda^{\gamma+1}/c_0^\gamma$. If I could observe search effort, these parameters could be separately identified. I elect to fix c_0 .

Moreover, while ρ could be estimated, there is literature that suggests it is difficult to do so (see, for example, Hall [1979] or Campbell et al. [1997], chap. 8). In light of this, I set the discount rate to the standard 5%. I find that varying this parameter between zero and 10% has no significant effect on the resulting estimates of other parameters.

3.4.3 Results

I coded a federal and provincial tax function, the aggregate of which is the income tax faced by individuals in Ontario. Then, I used the limited info MLE described above to obtain parameter estimates. Finally, subject to raising the *same* amount of tax revenue as under the current tax system, I solve for the optimal federal tax rates (keeping the thresholds the same). This involves maximizing over 4 marginal tax rate variables. As before, without incorporating any concept of inequality aversion, the optimal tax rate consists of raising all revenue from the lowest income individuals, and subsidizing (or setting tax rate = 0 if this is not possible) the high income individuals.

A simple fix to this problem is to introduce some concavity to the social welfare

function. In particular, I amend the government's objective function to:

$$\begin{aligned} \max_T \quad & uV_u^{1/\kappa} + (1-u) \int V_e(x)^{1/\kappa} g(x) dx \\ \text{s.t } & R(T) \geq c \end{aligned}$$

where κ represents the degree of inequality aversion of the social planner.

Solving the government's welfare maximizing problem subject to the "net-neutral" budget constraint (i.e., $c = 0$) and using $\kappa = 1.2$, I find that the federal tax rate should be set to 16.4% for the first \$43953 of taxable income, 5.9% for the next \$43954 of taxable income, 75.5% for the next \$48363 of taxable income, and 5.8% for taxable income over \$136270. Results are different for different choices of κ , but the same intuition persists: the government wants to subsidize search effort while generating adequate tax revenue. Since there is inequality aversion, the optimal strategy is no longer to solely tax the poorest individuals (since this would lead to high inequality), but to tax everyone *but* the richest, and generate more tax revenue from the middle-class since, due to inequality aversion, they will feel it less.

In other words, the government should lower taxes on the rich to incentive search, but raise taxes on the middle-class in order to raise its required revenue from the "right places", according to its inequality aversion, κ .

Table 3.3: Optimal Tax Rates using Ontario Subsample

	$\kappa = 1$	$\kappa = 1.2$	$\kappa = 1.5$	$\kappa = 2$
Tier 1	23.4	16.4	6.6	2.3
Tier 2	0	5.9	41.9	60.5
Tier 3	0	75.5	89.5	90.6
Tier 4	0	5.8	0.6	0

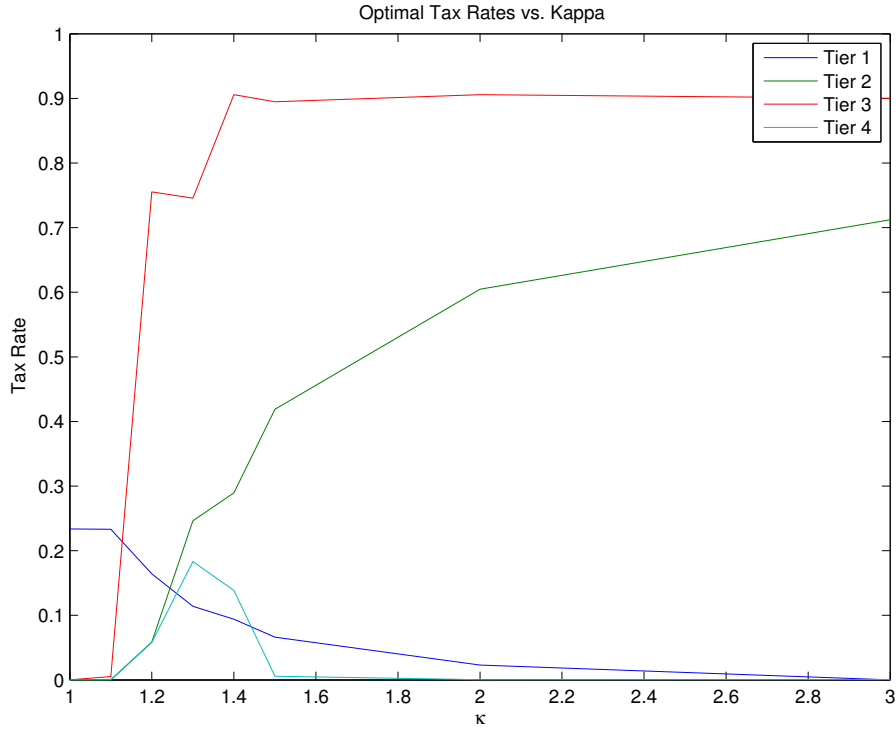


Figure 3.2: Optimal Federal Tax Rates vs. Inequality Aversion (Ontario)

The results are similar when I use data from all provinces in Canada - the only difference being the highest tax bracket marginal tax rate stays very close to zero. This is likely due to the more progressive provincial taxes in other provinces that already make the Tier 4 tax rates high (compare, for example, the highest marginal tax rate in Quebec at 26% vs. the highest marginal tax rate in Ontario at 13%).

Table 3.4: Optimal Tax Rates

	$\kappa = 1$	$\kappa = 1.3$	$\kappa = 1.9$	$\kappa = 2.5$
Tier 1	24.5	17.6	6.6	4
Tier 2	0	1.4	39.3	49
Tier 3	0	79.6	84.8	85.9
Tier 4	0	0.1	0	0

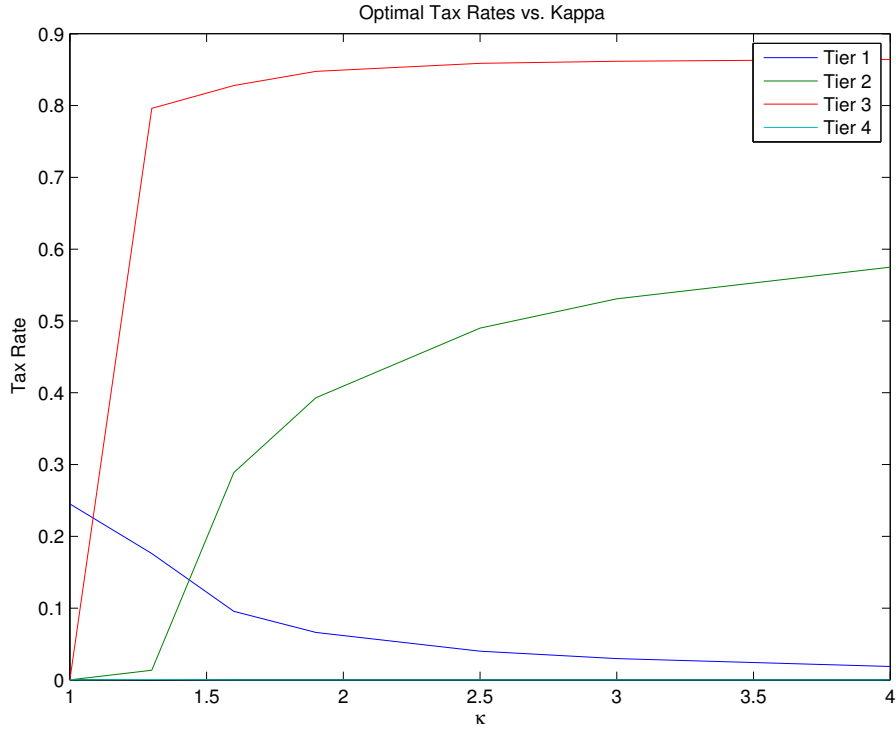


Figure 3.3: Optimal Federal Tax Rates vs. Inequality Aversion (All Provinces)

3.5 Conclusion

In this chapter, I extended Mirrlees' theory of optimal taxation (Mirrlees [1971]) to a framework with endogenous job search, and applied to the estimation of optimal federal tax rates in Canada. Consistent with previous literature, I find that the optimal federal

tax rates are zero for the highest income bracket. This is driven by endogenous search effort: the government wants to incentive workers to keep searching on the job until they receive and accept a higher wage. In general, I find that optimal tax rates are hump-shaped in income (i.e., they are smallest for very low and very high incomes), but the magnitude of the tax rates depends heavily on the level of inequality aversion the government has. In particular, for low levels of inequality aversion, it is optimal to tax the lowest bracket only, in order to induce them to increase search effort and obtain a higher wage; for high levels of inequality aversion, the optimal tax of the lowest bracket diminishes, and the government raises its revenue by taxing the middle two income brackets.

As an extension, one could estimate the general equilibrium model. I have outlined how to do so in the appendix. In effect, estimating this model amounts to solving a fixed-point problem in the wage offer distribution function, where each step requires solving two integral equations. This would allow firms to adjust the wages they choose to offer as the government changes incomes taxes, and may lead to different optimal tax predictions. One could also introduce corporate income taxes in this framework, and explore their interplay with individual income taxes. I leave such analysis for future work.

Bibliography

- M. Aguiar, E. Hurst, and L. Karabarbounis. The life-cycle profile of time spent on job search. *The American Economic Review*, 103(3):111–116, 2013.
- V. Aguirregabiria and C.-Y. Ho. A dynamic oligopoly game of the us airline industry: Estimation and policy experiments. *Journal of Econometrics*, 168(1):156–173, 2012.
- N. Aizawa and H. Fang. Equilibrium labor market search and health insurance reform. Technical report, National Bureau of Economic Research, 2013.
- G. Arie, S. Markovich, and M. J. Varela. The competitive effect of multimarket contact. 2014.
- J. Bagger and R. Lentz. An empirical model of wage dispersion with sorting. Technical report, National Bureau of Economic Research, 2014.
- P. Barla and C. Constantatos. Airline network structure under demand uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 36(3):173–180, 2000.
- C. Barnhart, A. M. Cohn, E. L. Johnson, D. Klabjan, G. L. Nemhauser, and P. H. Vance. Airline crew scheduling. In *Handbook of transportation science*, pages 517–560. Springer, 2003.
- C. L. Benkard, A. Bodoh-Creed, and J. Lazarev. Simulating the dynamic effects of horizontal mergers: Us airlines. *Manuscript, Yale University*, 2010.
- S. Berry and P. Jia. Tracing the woes: An empirical analysis of the airline industry. *American economic journal. Microeconomics*, 2(3):1–43, 2010.
- S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.
- S. Berry, M. Carnall, and P. T. Spiller. Airline hubs: costs, markups and the implications of customer heterogeneity. *Competition policy and antitrust*, 2006.
- S. T. Berry and P. A. Haile. Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Technical report, National Bureau of Economic Research, 2009.
- K. Burdett. A theory of employee job search and quit rates. *The American Economic Review*, 68(1):212–220, 1978.

- J. Y. Campbell, A. W.-C. Lo, A. C. MacKinlay, et al. *The econometrics of financial markets*, volume 2. princeton University press Princeton, NJ, 1997.
- S. Choi, A. Janiak, and B. Villena-Roldán. Unemployment, participation and worker flows over the life-cycle. *The Economic Journal*, 125(589):1705–1733, 2015.
- B. J. Christensen, R. Lentz, D. T. Mortensen, G. R. Neumann, and A. Werwatz. On-the-job search and the wage distribution. *Journal of Labor Economics*, 23(1):31–58, 2005.
- F. Ciliberto and J. W. Williams. Does multimarket contact facilitate tacit collusion? inference on conduct parameters in the airline industry. *The RAND Journal of Economics*, 45(4):764–791, 2014.
- F. Ciliberto, C. Murry, and E. T. Tamer. Market structure and competition in airline markets. *Available at SSRN 2777820*, 2016.
- J. Currie and B. C. Madrian. Health, health insurance and the labor market. *Handbook of labor economics*, 3:3309–3416, 1999.
- M. S. Dey and C. J. Flinn. An equilibrium model of health insurance provision and wage determination. *Econometrica*, 73(2):571–627, 2005.
- M. Draganska, M. Mazzeo, and K. Seim. Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *QME*, 7(2):105–146, 2009.
- A. Gandhi and J.-F. Houde. Measuring substitution patterns in differentiated products industries the missing instruments. Technical report, Mimeo, University of Wisconsin-Madison and University of Pennsylvania, 2015.
- D. W. Gillen, W. G. Morrison, and C. Stewart. Air travel demand elasticities: Concepts, issues and measurement. *Department of Finance, Government of Canada*, 2003.
- R. E. Hall. Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *NBER working paper*, (R0015), 1979.
- K. Ho. Insurer-provider networks in the medical care market. *The American Economic Review*, 99(1):393–430, 2009.
- T. J. Holmes. The diffusion of wal-mart and economies of density. *Econometrica*, 79(1):253–302, 2011.
- D. Hosken, N. Miller, and M. Weinberg. Ex post merger evaluation: How does it help ex ante? *Journal of European Competition Law & Practice*, 2016.
- J. Kwoka and E. Shumilkina. The price effect of eliminating potential competition: Evidence from an airline merger. *The journal of industrial economics*, 58(4):767–793, 2010.
- J. Lazarev. The welfare effects of intertemporal price discrimination: an empirical analysis of airline pricing in us monopoly markets. *New York University*, 2013.

- R. Lentz and D. T. Mortensen. Labor market models of worker and firm heterogeneity. *Annu. Rev. Econ.*, 2(1):577–602, 2010.
- C.-S. Li, C.-C. Liu, Y.-C. Kuo, and C.-S. Yang. Health insurance provision and labor contracts for small firms. *Small Business Economics*, 40(2):325–334, 2013.
- G. Li, I. Kwan, and D. Rutherford. Us domestic airline fuel efficiency ranking, 2014, 2015.
- D. Luo. The price effects of the delta/northwest airline merger. *Review of Industrial Organization*, 44(1):27–48, 2014.
- N. G. Mankiw, M. Weinzierl, and D. Yagan. Optimal taxation in theory and practice. *The Journal of Economic Perspectives*, 23(4):147–174, 2009.
- D. McFadden. Conditional logit analysis of qualitative choice behavior. 1973.
- J. A. Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208, 1971.
- D. T. Mortensen. Job search and labor market analysis. *Handbook of labor economics*, 2: 849–919, 1986.
- D. T. Mortensen and C. A. Pissarides. Job creation and job destruction in the theory of unemployment. *The review of economic studies*, 61(3):397–415, 1994.
- A. Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69 (2):307–342, 2001.
- A. Pakes, J. Porter, K. Ho, and J. Ishii. Moment inequalities and their application. Technical report, cemmap working paper, Centre for Microdata Methods and Practice, 2007.
- E. Pels. Airline network competition: Full-service airlines, low-cost airlines and long-haul markets. *Research in Transportation Economics*, 24(1):68–74, 2008.
- P. C. Reiss and P. T. Spiller. Competition and entry in small airline markets. *The Journal of Law & Economics*, 32(2):S179–S202, 1989.
- J. Roberts and A. Sweeting. The potential entry defense in airline mergers. Technical report, working paper, Duke University, 2014.
- N. G. Rupp and K. M. Tan. Mergers and product quality: A silver lining from de-hubbing in the us airline industry. 2016.
- E. Sadka. On income distribution, incentive effects and optimal income taxation. *The review of economic studies*, 43(2):261–267, 1976.
- E. Saez. Using elasticities to derive optimal income tax rates. *The review of economic studies*, 68(1):205–229, 2001.
- G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2): 461–464, 1978.

- J. K. Seade. On the shape of optimal tax schedules. *Journal of public Economics*, 7(2): 203–235, 1977.
- C. Shapiro and J. E. Stiglitz. Equilibrium unemployment as a worker discipline device. *The American Economic Review*, 74(3):433–444, 1984.
- Y. Wei. Network effects of air travel demand. 2014.
- K. R. Williams. Dynamic airline pricing and seat availability. Technical report, mimeo, 2013.

Appendices

A Appendix to Chapter 1

A.1 Supplementary Summary Statistics and Estimates

A.1.1 Summary Statistics for 2015Q3 Data

Below, I outline some summary statistics of the 2015Q3 sample used in estimation. I drop the same itineraries and markets as in the 2012Q3 sample. Here, I keep itineraries operated by the top 7 ticketing carriers, by passengers transported. These are, in order of passengers transported: Southwest, AA, Delta, United, JetBlue, Alaska, and Spirit. As before, I do not drop any itineraries by operating carrier.

After all of these restrictions, my sample contains 648 thousand products, and 4496 markets. This is comparable to the 2012Q3 sample. First, I present summary statistics on the demand-side observables that are aggregated at the itinerary-level. This is to say, the averages are not passenger-weighted, but “itinerary-weighted”. Notice that the average fare ends up being over \$500 this way. I define variables the same way as I did for the 2012Q3 sample. Notice that about 6% of links in my sample involve stopovers. This is less than half the percentage of links that had stopovers in the 2012Q3 sample.

Table A.1: Summary Statistics for the Observables (2015)

	Mean	SD	Min	Max
Connection	0.74	0.44	0	1
Fare (\$100s)	5.1	3.05	0.4	99
Tourist Cities	0.17	0.38	0	1
Vacation/Resort Locations	0.14	0.35	0	1
Number of Direct Flights	28.3	18.7	0	65
Distance (1000s miles)	2.61	1.48	0.13	11.8
Congested Ends	0.25	0.5	0	2
Congested Connections	0.26	0.58	0	2
Hub Origin	0.32	0.47	0	1
Number of Links w/ Stopovers	0.06	0.38	0	4

Below, I present some additional summary statistics, including the passenger-weighted fare. Notice how much lower the weighted fare is than the non-weighted one, at \$420. This means that, in our sample, the correlation between fare and number of passengers is negative. This suggests that larger groups of people are either better at finding lower fares (or more likely to get a group discount), or they tend to choose products with lower fares. The average number of products per market is 131; this is the same as in the 2012Q3 sample.

Table A.2: Other Summary Statistics (2015)

	Mean	SD	Min	Max
Products Per Market	144	181	1	1830
Unconditional Product Share (1e-5)	1.31	4.78	0.06	409
Conditional Product Share	0.01	0.02	0	1
Weighted Fare (\$100s)	4.2	2.56	0.5	100

Finally, I report summary statistics for the 7 ticketing carriers included in my sample.

Notice that Southwest offers *fewer* products than AA, but transports nearly 50% more passengers than AA. Southwest's growth from 2012 to 2015 is partially driven by the fact that it acquired AirTran Airways, which was, in 2012, the next largest carrier after JetBlue.

Table A.3: Summary Statistics for the Ticketing Carriers (2015)

	# Passengers	% Passengers	# Products	% Products
Southwest	914437	31.5	162512	25.1
American Airlines	645551	22.3	183761	28.4
Delta	557321	19.2	161663	25
United Airlines	400820	13.8	115809	17.9
JetBlue	177518	6.2	9071	1.5
Alaska Airlines	106869	3.7	8397	1.3
Spirit Airlines	102292	3.6	6463	1

A.1.2 Estimates using 2015Q3 Data with 2 AA Dummies

Table A.4: Demand-Side Parameter Estimates (2015)

	OLS (1-type)	IV (1-type)	GMM (1-type)	GMM (2-types)
Constant (Tourist)	-8.16*** (0.006)	-5.752*** (0.008)	-5.777*** (0.002)	-5.672*** (0)
Connection (Tourist)	-0.744*** (0.002)	-1.148*** (0.002)	-1.157*** (0.001)	-1.145*** (0)
Fare (Tourist)	-0.033*** (0)	-0.709*** (0.002)	-0.713*** (0.05)	-1.099*** (0.001)
Constant (Business)	- -	- -	- -	-7.65*** (0.001)
Connection (Business)	- -	- -	- -	-1.073*** (0.001)
Fare (Business)	- -	- -	- -	-0.305*** (0.035)
Stopovers	-0.263*** (0.004)	-0.369*** (0.004)	-0.372*** (0)	-0.369*** (0)
Tourist Cities	0.136*** (0.002)	0.258*** (0.002)	0.26*** (0)	0.259*** (0)
Vacation/Resort Locations	0.437*** (0.002)	0.118*** (0.002)	0.119*** (0)	0.119*** (0)
Direct Routes	0.007*** (0)	0.015*** (0)	0.015 (2.566)	0.015 (2.327)
Distance	0.035*** (0.002)	0.218*** (0.002)	0.215*** (0.015)	0.222*** (0.016)
Distance ²	-0.007*** (0)	0.025*** (0)	0.025 (0.214)	0.026 (0.221)
Congested Ends	-0.084*** (0.002)	-0.157*** (0.002)	-0.158*** (0)	-0.157*** (0)
Congested Connections	-0.03*** (0.001)	-0.047*** (0.001)	-0.047*** (0)	-0.047*** (0)
Hub Origin	-0.114*** (0.003)	-0.022*** (0.003)	-0.022*** (0)	-0.022*** (0)
Southwest	-0.389*** (0.005)	0.112*** (0.005)	0.112*** (0)	0.112*** (0)
Delta	-0.518*** (0.005)	0.723*** (0.006)	0.721*** (0)	0.731*** (0)
United	-0.579*** (0.005)	0.539*** (0.006)	0.54*** (0)	0.546*** (0)
American (Legacy)	-0.549*** (0.005)	0.394*** (0.006)	0.395*** (0)	0.397*** (0)
American (Acquired)	-0.546*** (0.005)	0.252*** (0.006)	0.254*** (0)	0.254*** (0)
JetBlue	-0.073*** (0.008)	0.474*** (0.008)	0.478*** (0)	0.477*** (0)
Lambda	0.535*** (0.001)	0.446*** (0.001)	0.444*** (0.077)	0.453*** (0.071)
Gamma	- -	- -	- -	0.597*** (0.003)
Mu	- -	- -	0.099 (0.184)	0.101 (0.177)
Test of Over Identification (p-value)	-	0	-	-
1st Stage R^2	-	0.32	-	-
1st Stage F-test	-	1363	-	-

Table A.5: Supply-Side Parameter Estimates (2015)

	OLS (no Cap Con)	OLS	IV
Short-Haul Constant	1.163*** (0.06)	1.181*** (0.06)	1.228*** (0.06)
Short-Haul Distance	0.317** (0.13)	0.304** (0.13)	0.445*** (0.13)
Short-Haul Distance ²	0.08 (0.06)	0.079 (0.06)	0.126** (0.06)
Short-Haul Connection	-0.169*** (0.01)	-0.173*** (0.01)	0.065*** (0.01)
Long-Haul Constant	0.408*** (0.06)	0.442*** (0.06)	0.361*** (0.06)
Long-Haul Distance	0.3*** (0.01)	0.268*** (0.01)	0.599*** (0.01)
Long-Haul Distance ²	0.031*** (0)	0.034*** (0)	0.009*** (0)
Long-Haul Connection	-0.532*** (0)	-0.542*** (0)	-0.343*** (0.01)
Hub (Origin, Dest or Transfer)	0.331*** (0.02)	0.322*** (0.02)	0.307*** (0.02)
Congested Ends	0.057*** (0.01)	0.054*** (0.01)	0.045*** (0.01)
Capacity Constrained	- (0)	0.051*** (0)	-0.472*** (0.01)
Southwest	0.545*** (0.03)	0.533*** (0.03)	0.521*** (0.03)
Delta	1.3*** (0.03)	1.291*** (0.03)	1.588*** (0.03)
United	1.33*** (0.03)	1.313*** (0.03)	1.548*** (0.03)
American (Legacy)	1.044*** (0.03)	1.03*** (0.03)	1.252*** (0.03)
American (Acquired)	0.833*** (0.03)	0.831*** (0.03)	1.199*** (0.03)
JetBlue	0.581*** (0.03)	0.604*** (0.03)	0.529*** (0.03)
Test of Over Identification (p-value)	-	-	0
1st Stage R^2	-	-	0.26
1st Stage F-test	-	-	643

A.2 Alternate Supply-Side Estimation Techniques

A.2.1 MC Regression

There is more information I can extract from the capacity constraint data. For example, I can make the marginal cost specification link-specific, as in Wei [2014], and introduce a link-level capacity constraint specification, perhaps depending on the nature of the link constrained (distance, centrality, connectedness, etc.) as opposed to an aggregated product-level constraint specification.

In other words, I can let the marginal cost on route j be

$$mc_j = \sum_{s:s \in j} g_s + \mu' w_j + \omega_j$$

where g_s is the marginal cost on link s , w_j is a vector of route-level characteristics that are cost-relevant but not captured in $g_s(\cdot)$, and ω_j is a route-specific fixed-effect that captures unobservable determinants of marginal cost.

Then, as in Berry et al. [2006], I can let:

$$g_s = \eta' w_s + h(d_s, F_s, \theta)$$

I can also vary the capacity constraint threshold, κ . Testing the resulting Bayesian Information Criterion (“BIC”; see Schwarz et al. [1978]) from 2SLS, I find that the “optimal” specification (where BIC is minimized) occurs when $\kappa = 0.97$. This results in about 1% of products having capacity constrained links. Below, I plot BIC and the % of products

with capacity constrained links against κ .

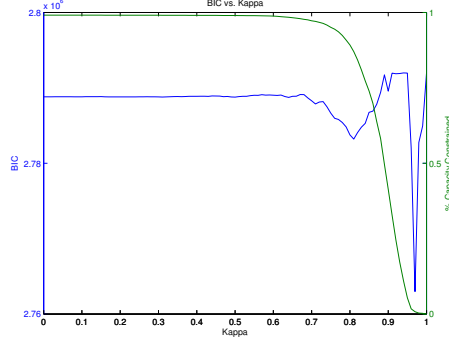


Figure A.1: BIC and % Products with a Capacity Constrained Link vs. κ

A.2.2 Structural Model

In this section, I discuss the issues with solving the full supply-side model. Solving this problem is equivalent to solving for the multipliers, μ . Since I already know B and ∇q , once I know μ , I can solve for marginal costs, and run the same regressions as before.

Usually, in a constrained problem, one can solve for the multipliers by using information from the constraints. The problem is that the constraints, as they are written, never bind. In other words, I never have that $B_{l,:}q = c_l$ for any link l . So I cannot use the condition that $B_{l,:}q = c_l$ for the capacity-constrained links in order to back-out μ_l .

Indeed, the empirical condition I have been using to generate “capacity-constrained” links has been $B_{l,:}q \geq \kappa$, where $\kappa < 1$. But I cannot use this either, since this constraint *is* violated by different products - some have load factors slightly above κ , while others have load factors much greater than κ .

One approach is to exploit the L additional equilibrium constraints from the aggregated load factors (where L is the number of links in my sample.) In particular, I know

that:

$$\begin{aligned}\text{Load}_j &= \frac{\text{Monthly Demand}_j}{\text{Monthly Capacity}_j} \\ &= \frac{Pr(D_j < c_j) * \mathbb{E}(D_j | D_j < c_j) + Pr(D_j \geq c_j) * c_j}{c_j}\end{aligned}$$

Where $D_j = \sum_{s:s \in j} D_s$, where s is the set of products that use the link j .

I can invert the above equations to obtain L additional conditions on prices. Since these will not correspond perfectly to the observed prices, if I can find a way to link the multipliers to prices in a well-grounded fashion, I can back-out the L multipliers.

One must ask then, what is the meaning of a multiplier if the constraint is regularly violated? It seems that an approach of this nature would necessarily require data on flight frequency (which is available but was not used in this paper), and a model that links individual capacity-constrained flights to an aggregate monthly load factor.

An alternate approach is to do away with this “discrete”, structural notion of capacity, in favor of a reduced-form continuous one. In particular, consider introducing a cost of being capacity-constrained into the firm’s optimization problem:

$$\max_p \pi_j = \sum_j (p_j - mc_j) q_j(p) - f(B, q(p), c)$$

This function could be non-zero only above a certain threshold, or even be discrete:

$$\max_p \pi_j = \sum_j (p_j - mc_j) q_j(p) - \rho \sum_l \mathbb{1}(B_l, q(p) \geq \kappa c_l)$$

I leave this problem for future work.

A.3 Additional Data

Since stopovers exist in coupon-level data, I cannot always know whether a coupon from point A to B with carrier C actually went there without stopping, or had a stopover at point S. If it's the former, I can use the segment capacity directly from the T-100, but if it's the latter, I have to first determine the intermediate airport S, and use information from the separate non-stop segments (A to S and S to B) to evaluate capacity and load from A to B. Finding S is non-trivial; indeed, it might be possible that the same carrier flies both a direct and non-stop flight between points A and B, and it may even fly multiple direct routes (e.g. A to S', S' to B)! This problem necessitates the use of airline-specific routing data, such as the On-Time Performance data.

The **On-Time Performance** database contains scheduled and actual departure and arrival times reported by certified US air carriers that account for at least one percent of domestic schedule passenger revenue. Moreover, it tells me the tail and flight number of every plane and flight on each non-stop segment. Using this data, I can map the routing of an individual plane across a month. I can also keep track of flight numbers, and determine the intermediate stopover cities, S, that an airplane may have taken on a direct flight between point A to B.

In order to match tail numbers to aircraft types, I need to use the publicly-available **FAA Aircraft Registry Database**. Matching requires two steps: 1) finding the aircraft manufacturer, model and series codes associated with each tail number (in the aircraft registration master file), and 2) finding the plane type and capacity corresponding to each unique manufacturer, model and series code (in the aircraft reference file). Note that I

can also find out the speed of each aircraft and other characteristics, from this matching process.

Since an airline may have multiple direct routes that it flies between point A and B (i.e., there may be multiple stopover combinations), it is difficult to match the non-stop segment T-100 load information with the coupon-level data of the DB1B. I leave this problem for future work.

A.4 Extensions

A.4.1 Multimarket Contact

I can incorporate tacit collusion in multimarket contact (“MMC”) by changing the 0 entries in Δ to be a function of multimarket contact, as in Ciliberto and Williams [2014]. Together with the capacity constraints, this model may be enough to empirically test the distinction between the competitive effect of MMC (C-MMC effect) and the mutual forbearance (tacit collusion) effect of MMC (MF-MMC effect) pioneered by Arie et al. [2014]. In other words, I can let:

$$\Delta_{jr} = \begin{cases} -\frac{\partial q_r}{\partial p_j} & \text{if } k = h \\ -f(mmc_{kh})\frac{\partial q_r}{\partial p_j} & \text{if } k \neq h \end{cases}$$

where

$$f(mmc_{kh}) = \frac{\exp(\phi_1 + \phi_2 mmc_{kh})}{1 + \exp(\phi_1 + \phi_2 mmc_{kh})} \in [0, 1]$$

A.4.2 Firm's Capacity Problem

Let an airline have k_i planes of type i , each with capacity s_i . Let the number of type i planes allocated to link l be given by n_{il} . Then, the first-stage problem can be written as:

$$\begin{aligned} \max_{n_{il}: l \in R} \quad & \pi(c^i, c^{-i}, \theta) \\ \text{s.t.} \quad & c_l = f(d_l) \sum_{i \in n} s_i n_{il} \\ & \sum_l n_{il} \leq k_i \end{aligned}$$

Where $f(d_l)$ captures the idea that planes on a shorter route can fly more frequently and, hence, deliver more capacity: $f'(d_l) < 0$.

Empirically, I see the types of planes allocated to each route, and can even follow individual planes (identified via their tail number) across trips. Then, I can solve for the fixed costs of routes with a moment inequality approach (see, for example, Pakes et al. [2007] and Holmes [2011]). I also leave this as an extension for future work.

A.4.3 Pricing Constraint

While this paper has largely been concerned with capacity as a network constraint, there is an additional network constraint that should not be ignored. In particular, consider the price of a trip with United from Philadelphia to London that has a layover in New York. If the price of the flight between Philadelphia and New York with United plus the price of the flight between New York and London were *less* than the price of the trip with

both of these bundled, an individual could easily (and at no additional cost⁷⁴) purchase the two round trip flights separately.

In other words, airlines are not free to choose whatever price they want for all of their products — indeed, as they lower the price of products, they could induce zero demand on other products they offer. They face a constrained maximization problem in prices.

The firm's problem would be re-written as follows:

$$\begin{aligned} \max_p \pi_j &= \sum_j (p_j - mc_j) q_j(p) \\ \text{s.t.} \quad p &\leq Ap \\ Bq(p) &\leq c \end{aligned}$$

Where the matrix A consists of 1's and 0's that capture the network constraint that the cost of a trip with a stop must not exceed the combined prices of the individual legs. Letting a_{ij} be the ij -th element of the matrix $A - I$, I get the following modifications to the firm's FOC condition:

$$q_j(p) + \sum_{r \in F_f} (p_r - mc_r) \frac{\partial q_r(p)}{\partial p_j} + \sum_l \lambda_l a_{lj} + \sum_l \mu_l \sum_k b_{lk} \frac{\partial q_k(p)}{\partial p_j} = 0$$

Re-writing the above in matrix form, and including the complementary slackness condi-

⁷⁴These arguments could be extended to include itineraries with multiple ticketing carriers, where there is some *cost* of switching carriers, capturing, say, the probability of a missed connection. The arguments above are for tickets with the *same* carrier, and it seems reasonable to assume that the disutility is negligible.

tions:

$$q - \Delta(p - mc) + (A - I)^T \lambda + [(\nabla q)^T B^T] \mu = 0$$

$$\lambda^T [(A - I)p] = 0$$

$$\mu^T [c - Bq] = 0$$

Then, as before, the first line can be re-written as:

$$mc = p - \Delta^{-1}(q - (A - I)^T \lambda - [(\nabla q)^T B^T] \mu)$$

Empirically, I will face a similar problem with pricing constraints. In particular, while I see a distribution of prices for routes, I do not know when tickets were purchased, or for what class of travel they were purchased for. One way to get around this is to assume a route, r , is **empirically price constrained** ($\lambda_r > 0$) if the ratio of the average monthly fare that route over its single-link-price components exceeds a threshold, $\kappa_2 \leq 1$:

$$\frac{p_r}{\sum_{j:j \in r} p_j} \geq \kappa_2$$

The question of how to know solve for these multipliers, λ , remains for future work.

B Appendix to Chapter 2

B.1 Lifecycle Job Search Model

Below, I outline the model and simulation results used to match the hump-shaped lifecycle profile of job search in the ATUS data identified by Aguiar et al. [2013]. I use the following three datasets:

1. CPS: Data from the Current Population Survey (CPS) is pooled to obtain unemployment rates and unemployment transition rates (as in Choi et al. [2015]). I use the 2013 Annual Social and Economic Supplement (ASEC) to obtain the percentage of workers with employer-sponsored health insurance.
2. ATUS (2003-2013): The American Time Use Survey (ATUS) is the main dataset that provides information on the time use of Americans - in particular, of a subset of CPS respondents. It details their daily schedules down to the minute. Unemployed job-searchers are unfortunately only a small subset of the ATUS respondents; over 10 years, I only have data on 6K individuals who engaged in any job search. Moreover, I cannot capture all the cohort effects present in this 10-year sample of respondents; yearly FE controls do not suffice.
3. MEPS (2011-2012, round 16): The Medical Expenditures Survey (MEPS) provides

information on *self-reported* health status and medical health expenditures. I divide individuals in this dataset into age bins, and obtain a distribution of medical health expenditures conditional on age and self-reported health status. In this way, I am able to estimate a health transition matrix.

The model can be summarized by the following value functions of unemployment, and employment:

$$V_{j,h}^u = \max_{c_j, s_j} u_h(c_j, 100 - s_j) + \beta \sum_{h'} Pr(h'|h, j) [\alpha(s_j) \int \max\{V_{j+1,h'}^u, V_{j+1,h'}^e(w)\} dF(w) + (1 - \alpha(s_j)) V_{j+1,h'}^u]$$

$$V_{j,h}^e(w) = \max_{c_j} u_h(c_j, 60) + \beta \sum_{h'} Pr(h'|h, w, j) [\delta V_{j+1,h'}^u + (1 - \delta) V_{j+1,h'}^e(w_{j+1})]$$

The following (separable) utility function:

$$u_h(c, l) = \frac{c^{1-\sigma_1}}{1-\sigma_2} + \psi \frac{l^{1-\sigma_2}}{1-\sigma_2} + h$$

The following search (production) function:

$$\alpha : [0, 100] \rightarrow [0, 1]$$

$$\alpha(s) = \left(\frac{s}{100}\right)^\kappa$$

And the associated first order conditions for optimal search effort.

$$\frac{u_2(c_j, l_j)}{\alpha'(s_j)} = \beta \mathbb{E}[\int \max\{V_{j+1}^e, V_{j+1}^u\} - V_{j+1}^u]$$

$$\frac{\partial s^*}{\partial (V^e - V^u)} > 0$$

If this (benchmark) asset accumulation model with endogenous search is not privy to any health considerations — i.e., health insurance is not tied to employment, then the resulting life-cycle profile of job search is decreasing, as evidenced in the other OECD countries.

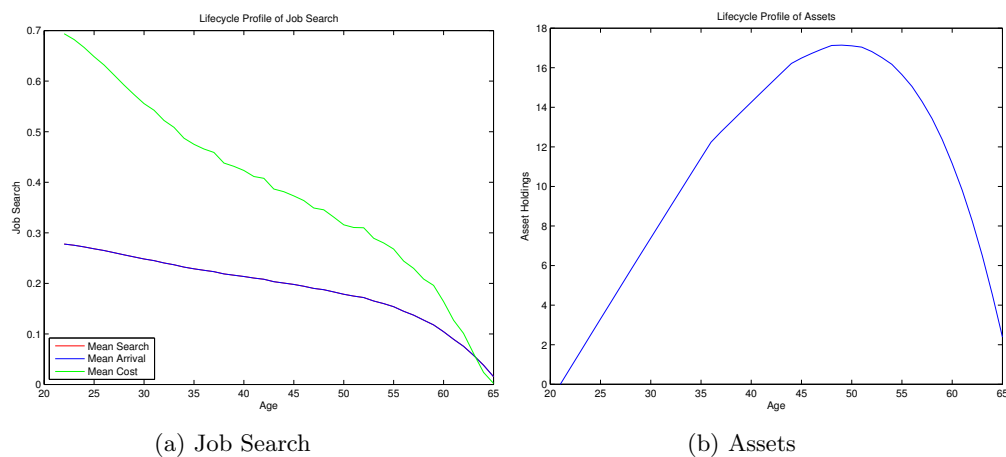


Figure B.1: Benchmark - No Health Considerations

On the other hand, if I introduce health insurance considerations - healthcare is tied to being employed, and if one is not insured they must bear the cost of a random draw from the estimated MEPS health expenditure distribution, conditional on health status that evolves according to the estimated health transition matrix - the resulting lifecycle profile of job search and hump-shaped and matches the ATUS data closely.

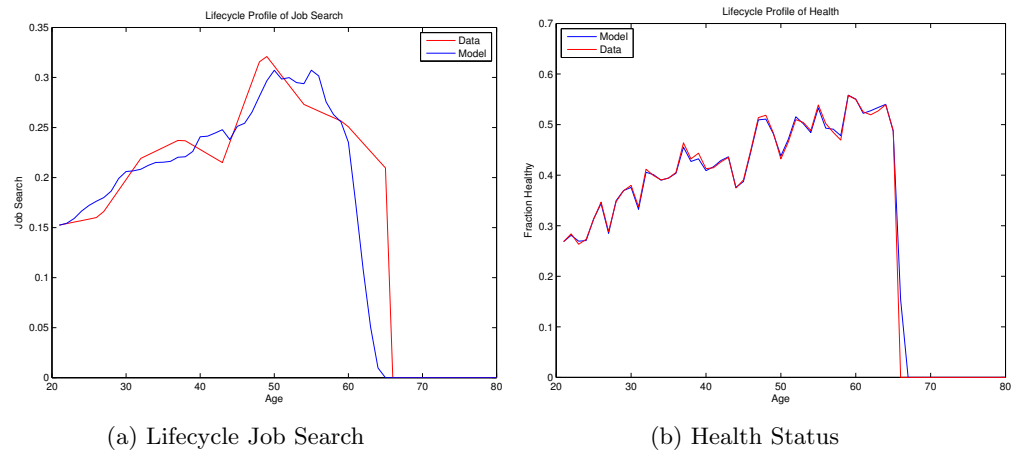


Figure B.2: MEPS Health Expenditure Data and Health Transitions

B.2 Proofs and Derivations

B.2.1 Proof of Certainty Equivalent given CARA Utility and Normal Shocks

I want to find $CE(x)$ such that $\mathbb{E}[-e^{-\gamma x}] = -e^{-\gamma CE(x)}$. Let $x \sim N(\bar{x}, V)$. Then,

$$\begin{aligned}
\mathbb{E}[-e^{-\gamma x}] &= \frac{1}{\sqrt{2\pi V}} \int -e^{-\gamma x} e^{-\frac{(x-\bar{x})^2}{2V}} dx \\
&= \frac{1}{\sqrt{2\pi V}} \int -e^{-\frac{2\gamma x V + (x-\bar{x})^2}{2V}} dx \\
&= \frac{1}{\sqrt{2\pi V}} \int -e^{-\frac{2\gamma x V + x^2 - 2x\bar{x} + \bar{x}^2 + (\gamma^2 V^2 - \gamma^2 V^2) + (-2\gamma V\bar{x} + 2\gamma V\bar{x})}{2V}} dx \\
&= \frac{1}{\sqrt{2\pi V}} \int -e^{-\frac{(x - (\bar{x} - \gamma V))^2 - \gamma^2 V^2 + 2\gamma\bar{x}V}{2V}} dx \\
&= -e^{-\frac{-\gamma^2 V + 2\gamma\bar{x}}{2}} \int \frac{1}{\sqrt{2\pi V}} e^{-\frac{(x - (\bar{x} - \gamma V))^2}{2V}} dx \\
&= -e^{-\frac{-\gamma^2 V + 2\gamma\bar{x}}{2}} \\
&= -e^{-\gamma(-\frac{1}{2}\gamma V + \bar{x})} \\
\Rightarrow CE(x) &= \bar{x} - \frac{1}{2}\gamma V
\end{aligned}$$

B.2.2 Proof of Theorem 1

Proof. If u is CARA, then it must necessarily be of the exponential form $u(x) = -e^{-\gamma x}$.

Then, letting $f_h(x)$ be the distribution of random medical expenditure shocks given health

status h ,

$$\begin{aligned}
x_h^{**} = 2 &\Rightarrow -e^{-\gamma[T(y)-R_h]} > \int -e^{-\gamma[T(y)-x]} f_h(x) dx \\
&\Rightarrow -e^{-\gamma[T(y)-R_h]} > -e^{-\gamma[T(y)]} \int e^{\gamma x} f_h(x) dx \\
&\Rightarrow e^{-\gamma[T(y)-R_h] + \gamma[T(y)]} < \int e^{\gamma x} f_h(x) dx \\
&\Rightarrow e^{\gamma R_h} < \int e^{\gamma x} f_h(x) dx \\
&\Rightarrow R_h < \frac{1}{\gamma} \log \int e^{\gamma x} f_h(x) dx \\
&\Rightarrow \xi_h < \frac{1}{\gamma \bar{m}_h} \log \int e^{\gamma x} f_h(x) dx - 1
\end{aligned}$$

Since this result is independent of y , substituting b for y delivers the same result, in other words:

$$x_h^* = 2 \Rightarrow e^{\gamma R_h} < \int e^{\gamma x} f_h(x) dx$$

This demonstrates equivalency between x_h^* and x_h^{**} .

Next, when medical expenditure shocks are normally distributed, I have that:

$$\begin{aligned}
x_h^{**} = 2 &\iff u(T(y) - R_h) > \mathbb{E}_{\tilde{m}_h} u(T(y) - \tilde{m}_h) = u(T(y) - \bar{m}_h - \frac{1}{2}\gamma \text{Var}(\tilde{m}_h)) \\
&\iff T(y) - R_h > T(y) - \bar{m}_h - \frac{1}{2}\gamma \text{Var}(\tilde{m}_h) \\
&\iff R_h < \bar{m}_h + \frac{1}{2}\gamma \text{Var}(\tilde{m}_h) \\
&\iff (1 + \xi_h)\bar{m}_h < \bar{m}_h + \frac{1}{2}\gamma \text{Var}(\tilde{m}_h) \\
&\iff 1 + \xi_h < 1 + \frac{1}{2}\gamma \frac{\text{Var}(\tilde{m}_h)}{\bar{m}_h} \\
&\iff \xi_h < \frac{1}{2}\gamma \frac{\text{Var}(\tilde{m}_h)}{\bar{m}_h}
\end{aligned}$$

Repeating the same exercise for x_h^* at $T(b)$ delivers the desired result. \square

B.2.3 Proof of Theorem 2

Proof. Let $CE(y, x)$ be the certainty equivalent of the gamble over the random medical shock x at the wealth level y , in other words: $\int u(y - x)f(x)dx = u(CE(y, x))$ (I omit the subscript h in this proof since this result holds independently for both health status types). Since u exhibits DARA, I have that $(y - CE(y, x))$ is decreasing in y . Then, since $T(y) > T(b)$ (by innocuous assumption), I have that $T(y) - CE(T(y), x) < T(b) - CE(T(b), x)$. Then the result immediately follows from:

$$\begin{aligned}
 x_h^{**} = 2 &\Rightarrow u(T(y) - R) > \int u(T(y) - x)f(x)dx = u(CE(T(y), x)) \\
 &\Rightarrow T(y) - R > CE(T(y), x) \\
 &\Rightarrow R < T(y) - CE(T(y), x) < T(b) - CE(T(b), x) \\
 &\Rightarrow x_h^* = 2
 \end{aligned}$$

\square

B.2.4 Derivation of Explicit Value Functions of Unemployment

Expanding,

$$U_h^y = (1 - \alpha)[v_h(b, x^*) + U_h^y] + \alpha \begin{cases} \frac{v_h(y, 1) - e + \frac{\delta}{1+r}[v_h(b, 1) + U_h^y]}{r + \delta}, & \text{if HI offered and NSC holds} \\ \frac{v_h(y, 1) + \frac{\delta + q}{1+r}[v_h(b, 1) + U_h^y]}{r + \delta + q}, & \text{if HI offered and NSC does not hold} \\ \frac{v_h(y, x^{**}) - e + \frac{\delta}{1+r}[v_h(b, x^*) + U_h^y]}{r + \delta}, & \text{if HI not offered and NSC holds} \\ \frac{v_h(y, x^{**}) + \frac{\delta + q}{1+r}[v_h(b, x^*) + U_h^y]}{r + \delta + q}, & \text{if HI not offered and NSC does not hold} \end{cases}$$

Simplifying,

$$U_h^y = \frac{1-\alpha}{\alpha} v_h(b, x^*) + \begin{cases} \frac{v_h(y, 1) - e + \frac{\delta}{1+r} [v_h(b, 1) + U_h^y]}{r + \delta}, & \text{if HI offered and NSC holds} \\ \frac{v_h(y, 1) + \frac{\delta+q}{1+r} [v_h(b, 1) + U_h^y]}{r + \delta + q}, & \text{if HI offered and NSC does not hold} \\ \frac{v_h(y, x^{**}) - e + \frac{\delta}{1+r} [v_h(b, x^*) + U_h^y]}{r + \delta}, & \text{if HI not offered and NSC holds} \\ \frac{v_h(y, x^{**}) + \frac{\delta+q}{1+r} [v_h(b, x^*) + U_h^y]}{r + \delta + q}, & \text{if HI not offered and NSC does not hold} \end{cases}$$

Now, notice that each case can be solved explicitly for U_h^y . For example, in the first case:

$$\begin{aligned} (r + \delta)U_h^y &= (r + \delta)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (v_h(y, 1) - e + \frac{\delta}{1+r} [v_h(b, 1) + U_h^y]) \\ [(1+r)(r + \delta) - \delta]U_h^y &= (1+r)(r + \delta)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, 1) - e) + \delta v_h(b, 1) \\ U_h^y &= \frac{(1+r)(r + \delta)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, 1) - e) + \delta v_h(b, 1)}{r(1 + \delta + r)} \end{aligned}$$

In the second case:

$$\begin{aligned} (r + \delta + q)U_h^y &= (r + \delta + q)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (v_h(y, 1) + \frac{\delta+q}{1+r} [v_h(b, 1) + U_h^y]) \\ [(1+r)(r + \delta + q) - (\delta + q)]U_h^y &= (1+r)(r + \delta + q)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, 1)) + (\delta + q)v_h(b, 1) \\ U_h^y &= \frac{(1+r)(r + \delta + q)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, 1)) + (\delta + q)v_h(b, 1)}{r(1 + \delta + r + q)} \end{aligned}$$

In the third case:

$$\begin{aligned} (r + \delta)U_h^y &= (r + \delta)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (v_h(y, x^{**}) - e + \frac{\delta}{1+r} [v_h(b, x^*) + U_h^y]) \\ [(1+r)(r + \delta) - \delta]U_h^y &= (1+r)(r + \delta)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, x^{**}) - e) + \delta v_h(b, x^*) \\ U_h^y &= \frac{(1+r)(r + \delta)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, x^{**}) - e) + \delta v_h(b, x^*)}{r(1 + \delta + r)} \end{aligned}$$

In the fourth case:

$$\begin{aligned}
(r + \delta + q)U_h^y &= (r + \delta + q)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (v_h(y, x^{**}) - e + \frac{\delta + q}{1+r}[v_h(b, x^*) + U_h^y]) \\
[(1+r)(r + \delta + q) - (\delta + q)]U_h^y &= (1+r)(r + \delta + q)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, x^{**}) - e) + (\delta + q)v_h(b, x^*) \\
U_h^{y_H^0} &= \frac{(1+r)(r + \delta + q)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, x^{**}) - e) + (\delta + q)v_h(b, x^*)}{r(1 + \delta + r + q)}
\end{aligned}$$

Notice that, throughout, the subscript h on the superscript y denotes the value function when the NSC does not hold for workers of health status h .

B.2.5 Derivation of No-Shirk-Conditions

Case 1: Firm offers HI

$$\begin{aligned}
V_h^H(1) &= V_h^L(1) \\
\frac{v_h(y, 1) - e + \delta U_h^{y_H^1}(1)}{r + \delta} &= \frac{v_h(y, 1) + (\delta + q)U_h^{y_H^1}(1)}{r + \delta + q} \\
qv_h(y, 1) &= (r + \delta)(\delta + q)U_h^{y_H^1}(1) - \delta(r + \delta + q)U_h^{y_H^1}(1) + (r + \delta + q)e \\
qv_h(y, 1) &= \frac{1}{1+r}[rqv_h(b, 1) + (r + \delta)(\delta + q)U_h^{y_H^1}(1) - \delta(r + \delta + q)U_h^{y_H^1}(1)] + (r + \delta + q)e \\
(1+r)qv_h(y, 1) &= rqv_h(b, 1) + (r + \delta)(\delta + q)U_h^{y_H^1}(1) - \delta(r + \delta + q)U_h^{y_H^1}(1) + (1+r)(r + \delta + q)e \\
(1+r)qv_h(y, 1) &= rqv_h(b, 1) + (r + \delta)(\delta + q)\left(\frac{(1+r)(r + \delta + q)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, 1) - e) + (\delta + q)v_h(b, 1)}{r(1 + \delta + r + q)}\right) - \\
&\quad - \delta(r + \delta + q)\left(\frac{(1+r)(r + \delta)\left(\frac{1-\alpha}{\alpha}\right)v_h(b, x^*) + (1+r)(v_h(y, 1) - e) + \delta v_h(b, 1)}{r(1 + \delta + r)}\right) + \\
&\quad + (1+r)(r + \delta + q)e
\end{aligned}$$

Where the fourth line follows from $U_h^{y_H^1}(1) = \frac{1}{1+r}[v_h(b, 1) + U_h^y]$.

Solving, (in Matlab), I get:

$$\begin{aligned}
0 &= \frac{(1+r)(q[v(y, 1) - v(b, 1) - \left(\frac{1-\alpha}{\alpha}\right)v(b, x^*)] - e(1 + \delta + q + r))}{r(1 + \delta + r)(1 + \delta + q + r)} \\
\Rightarrow v(y, 1) &= v(b, 1) + \left(\frac{1-\alpha}{\alpha}\right)v(b, x^*) + (1 + \delta + q + r)e/q
\end{aligned}$$

Case 2: Firm does not offer HI

$$\begin{aligned}
V_h^H(0) &= V_h^L(0) \\
\frac{v_h(y, x^{**}) - e + \delta U_h^{y_h^0}(x^*)}{r + \delta} &= \frac{v_h(y, x^{**}) + (\delta + q)U_h^{y_h^0}(x^*)}{r + \delta + q} \\
qv_h(y, x^{**}) &= (r + \delta)(\delta + q)U_h^{y_h^0}(x^*) - \delta(r + \delta + q)U_h^{y_h^0}(x^*) + (r + \delta + q)e \\
qv_h(y, x^{**}) &= \frac{1}{1 + r}[rqv_h(b, x^*) + (r + \delta)(\delta + q)U_h^{y_h^0} - \delta(r + \delta + q)U_h^{y_h^0}] + (r + \delta + q)e \\
(1 + r)qv_h(y, x^{**}) &= rqv_h(b, x^*) + (r + \delta)(\delta + q)U_h^{y_h^0} - \delta(r + \delta + q)U_h^{y_h^0} + (1 + r)(r + \delta + q)e \\
(1 + r)qv_h(y, x^{**}) &= rqv_h(b, x^*) + \\
&\quad + (r + \delta)(\delta + q) \frac{(1 + r)(r + \delta + q)(\frac{1 - \alpha}{\alpha})v_h(b, x^*) + (1 + r)(v_h(y, x^{**}) - e) + (\delta + q)v_h(b, x^*)}{r(1 + \delta + r + q)} - \\
&\quad - \delta(r + \delta + q) \left(\frac{(1 + r)(r + \delta)(\frac{1 - \alpha}{\alpha})v_h(b, x^*) + (1 + r)(v_h(y, x^{**}) - e) + \delta v_h(b, x^*)}{r(1 + \delta + r)} \right) + \\
&\quad + (1 + r)(r + \delta + q)e
\end{aligned}$$

Where the fourth line follows from $U_h^{y_h^0}(x^*) = \frac{1}{1 + r}[v_h(b, x^*) + U_h^y]$.

Solving, (in Matlab), I get:

$$\begin{aligned}
0 &= \frac{(1 + r)(q[v(y, x^{**}) - \frac{1}{\alpha}v(b, x^*)] - e(1 + \delta + q + r))}{r(1 + \delta + r)(1 + \delta + q + r)} \\
\Rightarrow v(y, x^{**}) &= \frac{1}{\alpha}v(b, x^*) + (1 + \delta + q + r)e/q
\end{aligned}$$

B.2.6 Proof of Theorem 4

Proof. From Theorem 2.1, I have that if u exhibits CARA, $x_h^* \iff x_h^{**}$. Then, subtracting the NSC of the good-health worker from the NSC of the bad-health worker, I can write:

$$v_g(y_g, x_g) - v_b(y_b, x_b) = \frac{1}{\alpha}[v_g(b, x_g) - v_b(b, x_b)] \quad (\text{B.2.1})$$

Where $x_h^{**} = x_h^* = x_h$.

Case 1: $x_b = x_g = 2$. Then, the equation above becomes:

$$\begin{aligned}
u(y_g - R_g) - u(y_b - R_b) &= \frac{1}{\alpha} [u(b - R_g) - u(b - R_b)] \\
-e^{-\gamma[y_g - R_g]} + e^{-\gamma[y_b - R_b]} &= \frac{1}{\alpha} [-e^{-\gamma[b - R_g]} + e^{-\gamma[b - R_b]}] \\
(-e^{\gamma R_g})(e^{-\gamma y_g}) + (e^{\gamma R_b})(e^{-\gamma y_b}) &= \frac{1}{\alpha} (e^{-\gamma b}) [-e^{\gamma R_g} + e^{\gamma R_b}]
\end{aligned}$$

Now consider, for a proof by contradiction, that $y_g \leq y_b$. Then:

$$\begin{aligned}
(-e^{\gamma R_g})(e^{-\gamma y_g}) + (e^{\gamma R_b})(e^{-\gamma y_b}) &\leq (-e^{\gamma R_g})(e^{-\gamma y_b}) + (e^{\gamma R_b})(e^{-\gamma y_b}) = e^{-\gamma y_b} [-e^{\gamma R_g} + e^{\gamma R_b}] \\
\Rightarrow \frac{1}{\alpha} (e^{-\gamma b}) [-e^{\gamma R_g} + e^{\gamma R_b}] &\leq e^{-\gamma y_b} [-e^{\gamma R_g} + e^{\gamma R_b}] \\
\Rightarrow \frac{1}{\alpha} (e^{-\gamma b}) &\leq e^{-\gamma y_b} \\
\Rightarrow e^{\gamma y_b} &\leq \alpha e^{\gamma b} \leq e^{\gamma b}
\end{aligned}$$

Where the third line follows since $R_g < R_b \Rightarrow [-e^{\gamma R_g} + e^{\gamma R_b}] > 0$. Since $\alpha \leq 1$, and $y_b > b$, I get the desired contradiction.

Case 2: $x_b = x_g = 0$. Then, the equation above becomes:

$$\begin{aligned}
u(y_g - \tilde{m}_g) - u(y_b - \tilde{m}_b) &= \frac{1}{\alpha} [u(b - \tilde{m}_g) - u(b - \tilde{m}_b)] \\
-\int e^{-\gamma[y_g - x_g]} f(x_g) dx_g + \int e^{-\gamma[y_b - x_b]} f(x_b) dx_b &= \frac{1}{\alpha} [-\int e^{-\gamma[b - x_g]} f(x_g) dx_g + \int e^{-\gamma[b - x_b]} f(x_b) dx_b] \\
(e^{-\gamma y_g})(\int -e^{\gamma x_g} f(x_g) dx_g) + (e^{-\gamma y_b})(\int e^{\gamma x_b} f(x_b) dx_b) &= \frac{1}{\alpha} (e^{-\gamma b}) [-\int e^{\gamma x_g} f(x_g) dx_g + \int e^{\gamma x_b} f(x_b) dx_b]
\end{aligned}$$

Again consider, for a proof by contradiction, that $y_g \leq y_b$. Then:

$$\begin{aligned}
(e^{-\gamma y_g})(\int -e^{\gamma x_g} f(x_g) dx_g) + (e^{-\gamma y_b})(\int e^{\gamma x_b} f(x_b) dx_b) &\leq (e^{-\gamma y_b})(\int -e^{\gamma x_g} f(x_g) dx_g) + (e^{-\gamma y_b})(\int e^{\gamma x_b} f(x_b) dx_b) \\
&= e^{-\gamma y_b}[-\int e^{\gamma x_g} f(x_g) dx_g + \int e^{\gamma x_b} f(x_b) dx_b] \\
\Rightarrow \frac{1}{\alpha}(e^{-\gamma b})[-\int e^{\gamma x_g} f(x_g) dx_g + \int e^{\gamma x_b} f(x_b) dx_b] &\leq e^{-\gamma y_b}[-\int e^{\gamma x_g} f(x_g) dx_g + \int e^{\gamma x_b} f(x_b) dx_b] \\
&\Rightarrow \frac{1}{\alpha}(e^{-\gamma b}) \leq e^{-\gamma y_b} \\
&\Rightarrow e^{\gamma y_b} \leq \alpha e^{\gamma b} \leq e^{\gamma b}
\end{aligned}$$

Where the fourth line follows since \tilde{m}_b first order stochastically dominates \tilde{m}_g , which implies that $F_b(x) < F_g(x) \forall x$ (with a strict inequality for some x) and, hence, that $\int e^{\gamma x_b} f(x_b) dx_b > \int e^{\gamma x_g} f(x_g) dx_g$. Since $\alpha \leq 1$, and $y_b > b$, I get the desired contradiction.

Case 3: $x_b = 2$ and $x_g = 0$. Then, the equation above becomes:

$$\begin{aligned}
u(y_g - \tilde{m}_g) - u(y_b - R_b) &= \frac{1}{\alpha}[u(b - \tilde{m}_g) - u(b - R_b)] \\
- \int e^{-\gamma[y_g - x_g]} f(x_g) dx_g + e^{-\gamma[y_b - R_b]} &= \frac{1}{\alpha}[- \int e^{-\gamma[b - x_g]} f(x_g) dx_g + e^{-\gamma[b - R_b]}] \\
(e^{-\gamma y_g})(- \int e^{\gamma x_g} f(x_g) dx_g) + (e^{-\gamma y_b})(e^{\gamma R_b}) &= \frac{1}{\alpha}(e^{-\gamma b})[- \int e^{\gamma x_g} f(x_g) dx_g + e^{\gamma R_b}]
\end{aligned}$$

Again consider, for a proof by contradiction, that $y_g \leq y_b$. Then:

$$\begin{aligned}
(e^{-\gamma y_g})(- \int e^{\gamma x_g} f(x_g) dx_g) + (e^{-\gamma y_b})(e^{\gamma R_b}) &\leq (e^{-\gamma y_b})(- \int e^{\gamma x_g} f(x_g) dx_g) + (e^{-\gamma y_b})(e^{\gamma R_b}) \\
&= e^{-\gamma y_b}[- \int e^{\gamma x_g} f(x_g) dx_g + e^{\gamma R_b}] \\
\Rightarrow \frac{1}{\alpha}(e^{-\gamma b})[- \int e^{\gamma x_g} f(x_g) dx_g + e^{\gamma R_b}] &\leq e^{-\gamma y_b}[- \int e^{\gamma x_g} f(x_g) dx_g + e^{\gamma R_b}] \\
&\Rightarrow \frac{1}{\alpha}(e^{-\gamma b}) \leq e^{-\gamma y_b} \\
&\Rightarrow e^{\gamma y_b} \leq \alpha e^{\gamma b} \leq e^{\gamma b}
\end{aligned}$$

Where the fourth line follows since:

$$\begin{aligned}
x_g = 0 &\Rightarrow - \int e^{-\gamma[y_g - x_g]} f(x_g) dx_g > -e^{-\gamma[y_g - R_g]} \\
e^{-\gamma y_g} \int e^{\gamma x_g} f(x_g) dx_g &< e^{-\gamma y_g} e^{\gamma R_g} \\
\int e^{\gamma x_g} f(x_g) dx_g &< e^{\gamma R_g} < e^{\gamma R_b} \\
\Rightarrow - \int e^{\gamma x_g} f(x_g) dx_g + e^{\gamma R_b} &> 0
\end{aligned}$$

Since $\alpha \leq 1$, and $y_b > b$, I get the desired contradiction.

Case 4: $x_b = 0$ and $x_g = 2$. Then, the equation above becomes:

$$\begin{aligned}
u(y_g - R_g) - u(y_b - \tilde{m}_b) &= \frac{1}{\alpha} [u(b - R_g) - u(b - \tilde{m}_b)] \\
-e^{-\gamma[y_g - R_g]} + \int e^{-\gamma[y_b - x_b]} f(x_b) dx_b &= \frac{1}{\alpha} [-e^{-\gamma[b - R_g]} + \int e^{-\gamma[b - x_b]} f(x_b) dx_b] \\
(e^{-\gamma y_g})(-e^{\gamma R_g}) + (e^{-\gamma y_b})(\int e^{\gamma x_b} f(x_b) dx_b) &= \frac{1}{\alpha} (e^{-\gamma b}) [-e^{\gamma R_g} + \int e^{\gamma x_b} f(x_b) dx_b]
\end{aligned}$$

Again consider, for a proof by contradiction, that $y_g \leq y_b$. Then:

$$\begin{aligned}
(e^{-\gamma y_g})(-e^{\gamma R_g}) + (e^{-\gamma y_b})(\int e^{\gamma x_b} f(x_b) dx_b) &\leq (e^{-\gamma y_b})(-e^{\gamma R_g}) + (e^{-\gamma y_b})(\int e^{\gamma x_b} f(x_b) dx_b) \\
&= e^{-\gamma y_b} [-e^{\gamma R_g} + \int e^{\gamma x_b} f(x_b) dx_b] \\
\Rightarrow \frac{1}{\alpha} (e^{-\gamma b}) [-e^{\gamma R_g} + \int e^{\gamma x_b} f(x_b) dx_b] &\leq e^{-\gamma y_b} [-e^{\gamma R_g} + \int e^{\gamma x_b} f(x_b) dx_b] \\
\Rightarrow \frac{1}{\alpha} (e^{-\gamma b}) &\leq e^{-\gamma y_b} \\
\Rightarrow e^{\gamma y_b} &\leq \alpha e^{\gamma b} \leq e^{\gamma b}
\end{aligned}$$

Where the fourth line follows since:

$$\begin{aligned} x_g = 2 &\Rightarrow e^{\gamma R_g} < \int e^{\gamma x_g} f(x_g) dx_g < \int e^{\gamma x_b} f(x_b) dx_b \\ &\Rightarrow -e^{\gamma R_g} + \int e^{\gamma x_b} f(x_b) dx_b > 0 \end{aligned}$$

Due to first-order stochastic dominance. Since $\alpha \leq 1$, and $y_b > b$, I get the desired contradiction. \square

B.2.7 Proof of Theorem 5

Proof. From Theorem 1, I have that if u exhibits CARA, $x_h^* \iff x_h^{**}$.

Case 1: $x_h^* = x_h^{**} = 2$. Then, the above equality becomes $u(y^1) - u(y^0 - R) = u(b) - u(b - R)$. Then:

$$\begin{aligned} -e^{-\gamma y^1} + e^{-\gamma[y^0 - R]} &= -e^{-\gamma b} + e^{-\gamma[b - R]} \\ e^{\gamma R}[e^{-\gamma y^0} - e^{-\gamma b}] &= e^{-\gamma y^1} - e^{-\gamma b} \\ e^{-\gamma y^0} - e^{-\gamma b} &> e^{-\gamma y^1} - e^{-\gamma b} \\ -\gamma y^0 &> -\gamma y^1 \\ y^0 &< y^1 \end{aligned}$$

Where the third line follows from $e^{\gamma R} > 1$ and $[e^{-\gamma y^0} - e^{-\gamma b}] < 0$ since $y^0 > b$.

Case 2: $x_h^* = x_h^{**} = 0$. Then, the above equality becomes $u(y^1) - \mathbb{E}u(y^0 - \tilde{m}) =$

$u(b) - \mathbb{E}u(b - \tilde{m})$. Then:

$$\begin{aligned} -e^{-\gamma y^1} + \int e^{-\gamma[y^0-x]} f(x) dx &= -e^{-\gamma b} + \int e^{-\gamma[b-x]} f(x) dx \\ -e^{-\gamma y^1} + e^{-\gamma y^0} \int e^{-\gamma[-x]} f(x) dx &= -e^{-\gamma b} + e^{-\gamma b} \int e^{-\gamma[-x]} f(x) dx \end{aligned}$$

Let $\int e^{\gamma x} f(x) dx = Q$. Notice that since $x \geq 0$ (medical shocks cannot be negative), and there exist some x such that $x > 0$ (since $\bar{m}_h > 0$), I must have that $Q > 1$. Then the proof is as in Case 1:

$$\begin{aligned} -e^{-\gamma y^1} + e^{-\gamma y^0} Q &= -e^{-\gamma b} + e^{-\gamma b} Q \\ Q[e^{-\gamma y^0} - e^{-\gamma b}] &= e^{-\gamma y^1} - e^{-\gamma b} \\ y^0 &< y^1 \end{aligned}$$

□

B.2.8 Proof of Theorem 6

Proof. From Theorem 2.2, I have that if u exhibits DARA, $x_h^{**} \Rightarrow x_h^*$.

Case 1: $x_h^* = x_h^{**} = 2$. Then, the above equality becomes $u(y^1) - u(y^0 - R) = u(b) - u(b - R)$. Then, by concavity of u and $y^1 > b$, I must have that $y^1 > y^0$.

Case 2: $x_h^* = x_h^{**} = 0$. Then, the above equality becomes $u(y^1) - \mathbb{E}u(y^0 - \tilde{m}) = u(b) - \mathbb{E}u(b - \tilde{m})$. As before, let $CE(y, x)$ denote the certainty equivalent of the no-insurance gamble over the random medial expenditure shocks x at the wealth level y .

Then:

$$u(y^1) - u(CE(y^0, x)) = u(b) - u(CE(b, m)) \quad (\text{B.2.2})$$

For a proof by contradiction, assume that $y^1 \leq y^0$. Then:

$$\begin{aligned} u(b) - u(CE(b, m)) &= u(y^1) - u(CE(y^0, x)) \\ &\leq u(y^0) - u(CE(y^0, x)) \end{aligned}$$

Since u is DARA, I have that $y^0 - CE(y^0, x) < b - CE(b, x)$. Concavity of u and $y^0 > b$ delivers the desired contradiction.

Case 3: $x_h^* = 2$ and $x_h^{**} = 0$. Then, the above equality becomes $u(y^1) - \mathbb{E}u(y^0 - \tilde{m}) = u(b) - u(b - R)$. Since $x_h^{**} = 2 \Rightarrow \mathbb{E}u(y^0 - \tilde{m}) > u(y^0 - R)$, I have that $u(y^1) - u(y^0 - R) > u(y^1) - \mathbb{E}u(y^0 - \tilde{m})$, so $u(y^1) - u(y^0 - R) > u(b) - u(b - R)$. Then, the proof is the same as in Case 1. \square

B.2.9 Proof of Theorem 14

Proof. Expanding (2.3.19), I get:

$$u(y^1) - \max\{u(y^0 + S(y^0) - R), \mathbb{E}u(y^0 - m_h - P(y))\} = u(b) - \max\{u(b + S(b) - R), \mathbb{E}u(b - m_h - P(b))\}$$

Exploiting linear utility as a consequence of risk-neutrality, I get:

$$\begin{aligned} y^1 - y^0 - \max\{S(y^0) - R, -\bar{m}_h - P(y^0)\} &= -\max\{S(b) - R, -\bar{m}_h - P(b)\} \\ y^1 - y^0 &= \max\{S(y^0) - R, -\bar{m}_h - P(y^0)\} - \max\{S(b) - R, -\bar{m}_h - P(b)\} \\ y^0 - y^1 &= \max\{R - S(y^0), \bar{m}_h + P(y^0)\} - \max\{R - S(b), \bar{m}_h + P(b)\} \end{aligned}$$

Since $S(y^0) < S(b)$ and $P(y^0) > P(b)$ (since $y^0 > b$), I have that $R - S(y^0) > R - S(b)$ and $\bar{m}_h + P(y^0) > \bar{m}_h + P(b)$. Since the max operator is just an upper envelope, it preserves the inequality. In other words, $y^0 - y^1 > 0$. This completes the proof. \square

B.2.10 Proof of Theorem 15

Proof. From Theorem 1, I have that:

$$x_h^{*B} = 2 \iff x_h^{**B} = 2 \iff R_h^B < \frac{1}{\gamma} \log \int e^{\gamma x} f_h(x) dx$$

And from Theorem 8, I have that:

$$\begin{aligned} x_h^{*ACA} = 2 &\iff R^{ACA} < \frac{1}{\gamma} \log \int e^{\gamma x} f_h(x) dx + S(b) + P(b) \\ x_h^{**ACA} = 2 &\iff R^{ACA} < \frac{1}{\gamma} \log \int e^{\gamma x} f_h(x) dx + S(y) + P(y) \end{aligned}$$

Then,

$$\begin{aligned} x_b^{*B} = 2 &\Rightarrow R_b^B < \frac{1}{\gamma} \log \int e^{\gamma x} f_b(x) dx \\ &\Rightarrow R^{ACA} < \frac{1}{\gamma} \log \int e^{\gamma x} f_b(x) dx \\ &\Rightarrow R^{ACA} < \frac{1}{\gamma} \log \int e^{\gamma x} f_b(x) dx + S(b) + P(b) \\ &\Rightarrow x_b^{*ACA} = 2 \end{aligned}$$

I can show the same for x_b^{**ACA} . To prove the second line,

$$\begin{aligned}
x_b^{*ACA} = 0 &\Rightarrow R^{ACA} > \frac{1}{\gamma} \log \int e^{\gamma x} f_b(x) dx + S(b) + P(b) \\
&\Rightarrow R^{ACA} > \frac{1}{\gamma} \log \int e^{\gamma x} f_b(x) dx \\
&\Rightarrow R_b^B > \frac{1}{\gamma} \log \int e^{\gamma x} f_b(x) dx \\
&\Rightarrow x_b^{*B} = 0
\end{aligned}$$

Finally, from Theorem 10, I have that $x_b^{*ACA} = 0 \Rightarrow x_g^{*ACA} = 0$. Analogous proof follows for the third line, at x_b^{**ACA} . □

B.3 Simulation

In this section, I simulate how changes in the exogenous variables $P_W(y)$, $S(y, R^{EX})$, $T(y)$, $P_E(n)$, etc.⁷⁵ will affect the equilibrium offer of wage and health insurance by the representative firm as well as the decision to self-insure by individuals. I hope to establish cut-off regions where firms will change their choice of offering health insurance, or firms will stop ensuring that certain workers exert high effort on the job, or individuals change their self-insurance decision.⁷⁶

I impose the following functional forms: $S(y, R^{EX}) = k_2 R^{EX} - k_3 y$, $P_W(y) = k_1 y$, $T(y) = y$, and $P_E(n) = K$ where K is a constant. I can later make $K(n)$, where n would represent the number of employed workers in the steady-state, $n = n_g + n_b$.

⁷⁵Notice that m_h is just the expected medical expenditure given to us by the data, firms do not enter into the health insurance exchange market, they just absorb the costs of the employees they provide insurance to, in a risk-neutral fashion.

⁷⁶As an extension for future work, one could also make q endogenous after introducing a cost-of-monitoring function $f(q)$, or make $q(e)$ and endogenize cost-of-effort e . Moreover, one could let health status change over time, introduce firm heterogeneity, make $b(y)$ or make COBRA continuation not occur after firing due to being caught shirking (with probability q).

B.3.1 Firm Offers HI: Support for Theorem 12

In this subsection, I abstract from the firm's decision to provide HI by focusing solely on the subcase where it does, indeed, offer HI. The simulations below support Theorem 12:

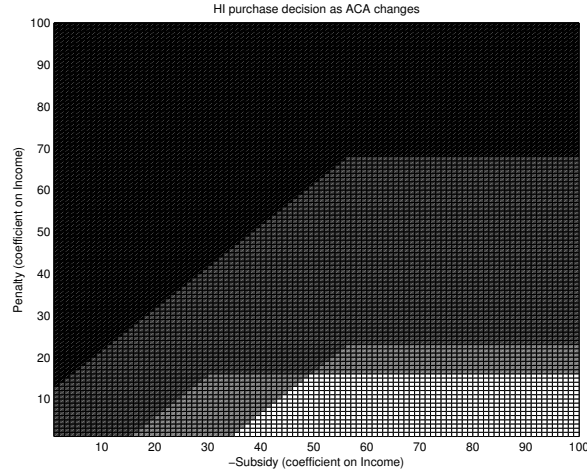


Figure B.3: HI Offered: ACA Effects on HI Coverage while Unemployed

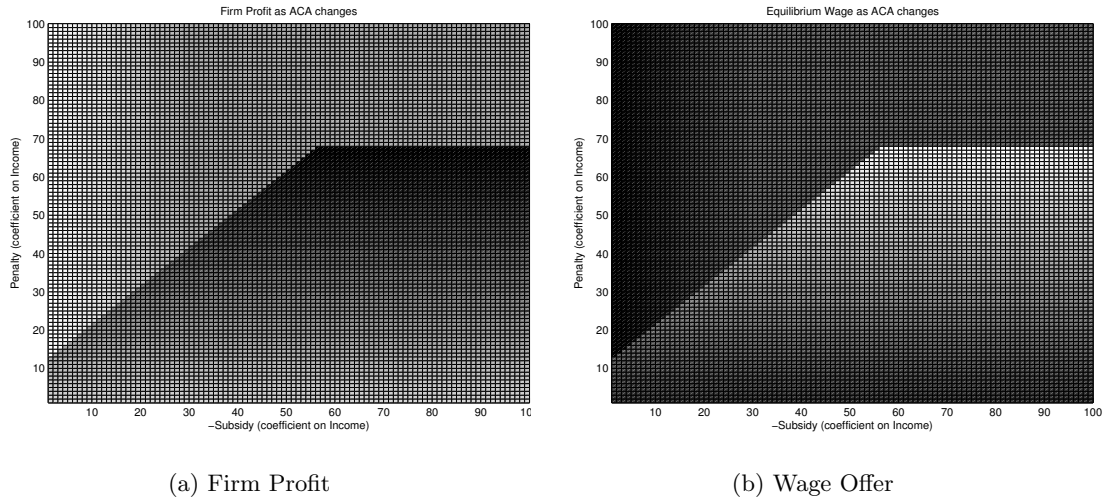


Figure B.4: HI Offered: ACA Effects on Profits and Wage

The darker the color, the higher the wage, profit or case number. I denote a special case 2.5 where both cases 2 and 3 can be supported in equilibrium-notice that it is in

this area that the firm can exploit multiple equilibria to choose the “bad” equilibrium for individuals (such that the good health remain uninsured), at a lower equilibrium wage and, consequently, higher profits.

What drives this? Notice first that I proved that $y_g^1 > y_b^1$ —this is to say, it is the good health workers’ NSC that is binding. Consequently, when it becomes optimal for them not to purchase HI, the firm can get away with offering them a lower wage (along with health insurance), knowing that this health insurance offer is now more available since I are stuck in the “bad” equilibrium where only the bad-health workers buy health insurance (recall that this drives up the price of health insurance R^{EX}).

B.3.2 Complete Firm Decision:

Next, I take into account the firm’s decision regarding whether or not to provide HI (with the stipulation that it faces a penalty if it chooses not to provide HI). Note that the simulations below use slightly different parameter configurations than the simulations in the previous subsection.

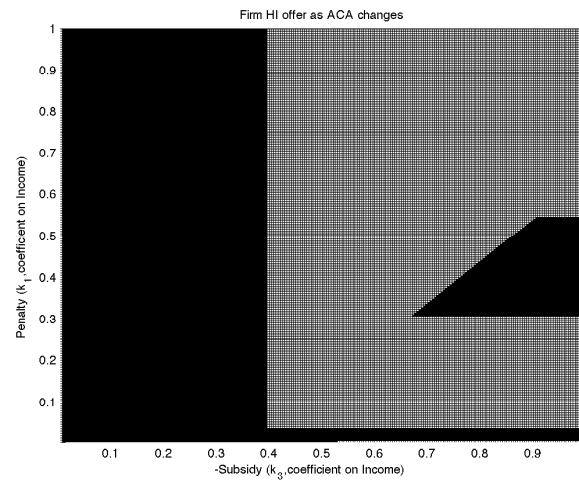


Figure B.5: Complete Decision: ACA Effects on Firm's Choice of HI offer

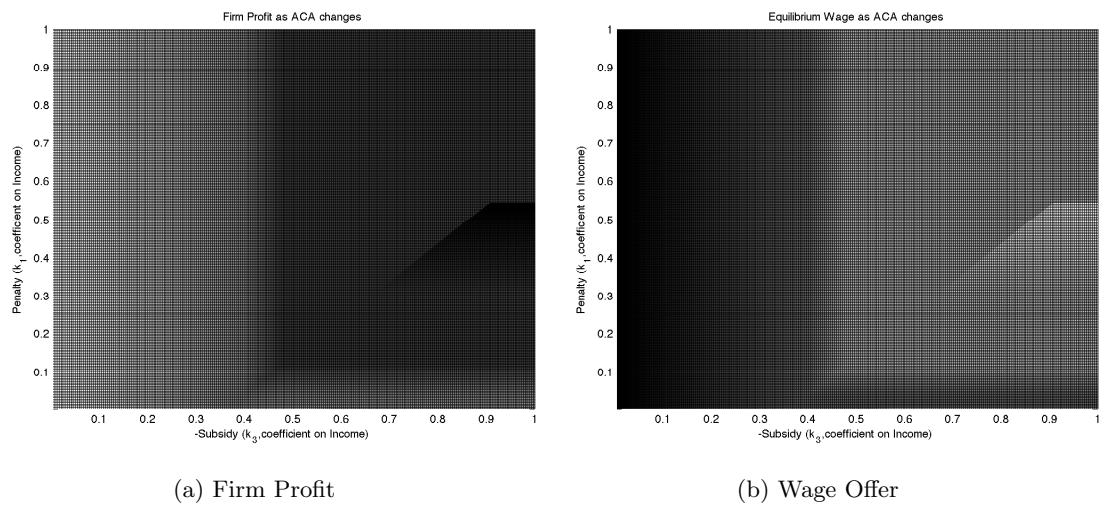
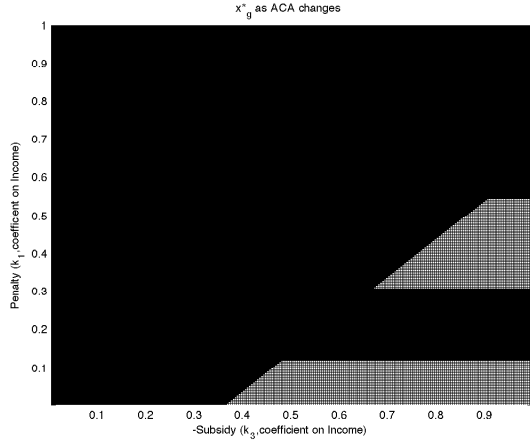
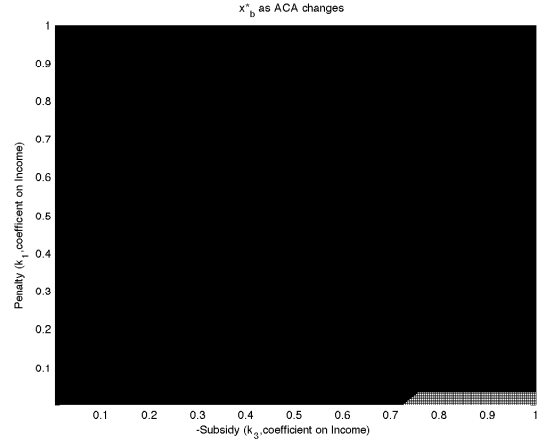


Figure B.6: Complete Decision: ACA Effects on Profits and Wage

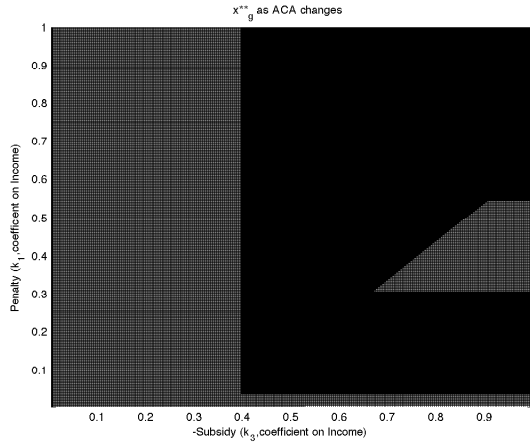


(a) Good Health Self-Insurance Decision

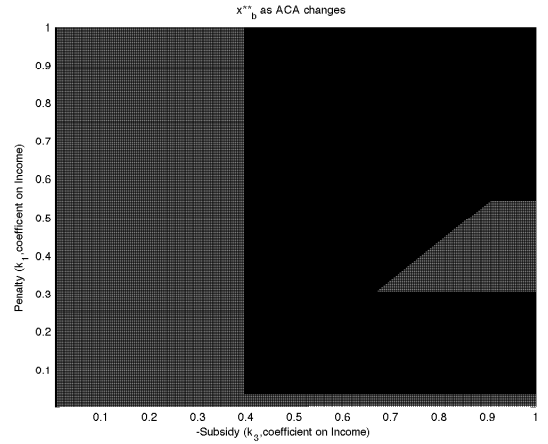


(b) Bad Health Self-Insurance Decision

Figure B.7: Complete Decision: ACA Effects on Self-Insurance Decision of Unemployed Workers



(a) Good Health Self-Insurance Decision



(b) Bad Health Self-Insurance Decision

Figure B.8: Complete Decision: ACA Effects on Self-Insurance Decision of Employed Workers

In the plot describing the firm's choice of whether or not to offer HI, black indicates that the firm offers HI while white indicates that the firm does not offer HI. In the self-insurance decision plots, black indicates that the individual self-insures, grey indicates that the individual receives HI from their firm, and white indicates that the individual

remains uninsured.

Notice that as subsidies increase (moving right to left on the x-axis), the firm offers health insurance, earns smaller profits, and offers a higher equilibrium wage—which is consistent with some of the theorems proved in the main body of the paper. The firm’s offer of HI can be explained as follows: when subsidies are very high, the firm offering HI does not have so much of a “double-edged sword” effect discussed above, since incentives to shirk and maintain existing HI for an additional period (due to COBRA continuation payments) is diminished.

The existence of such regions supports the idea that the government can induce higher wages by increasing premium subsidies; such subsidies induce firms to offer HI and individuals to self-insure, and help satisfy the “effective premiums under the ACA lower than in the benchmark” condition required for my wage-increase results. I add an additional plot below to further support higher wages under the ACA:

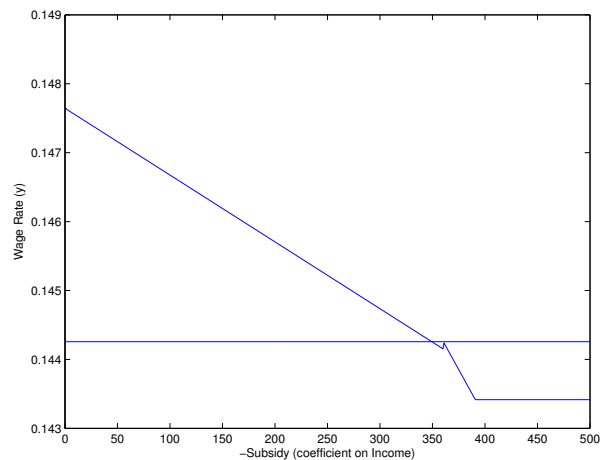


Figure B.9: Wage as Subsidies Decrease, Plotted Against Constant Benchmark Wage

Notice how wages are higher under the ACA above a certain threshold of premium

subsidies (subsidies increase right to left), as predicted by my Efficiency Wage Theorems (the small spike in wage occurs when the firm goes from offering HI to not offering HI).

In other words, here is the “cheap” intuition for the Efficiency Wage Result: as subsidies increases, employers are more inclined to offer HI to their workers. In turn, this changes the composition of health status of individuals using the insurance exchanges, driving premiums down. In turn, the relative cost of unemployment goes down, forcing firms to increase wages in order to induce good (no-shirk) behavior.

C Appendix to Chapter 3

C.1 General Equilibrium

C.1.1 Worker's Problem

As an extension, consider what happens if the firm's wage-offer function is endogenous.

Since no firm will offer a wage below w_R (it will be rejected for sure), I have that $\bar{F}(w_R) =$

1. As a result, the steady-state distribution of unemployed workers will be:

$$\begin{aligned} u &= \frac{\delta}{\delta + \lambda s^* \bar{F}(w_R)} \\ &= \frac{\lambda}{\delta + \lambda s^*} \end{aligned}$$

And

$$\frac{u}{1 - u} = \frac{\delta}{\lambda s^*} \tag{C.1.1}$$

The steady-state distribution of employed workers will be given by the implicit equa-

tion:

$$\begin{aligned}\lambda u s^*[F(w) - F(w_R)] &= \delta(1 - u)G(w) + \lambda(1 - u)\bar{F}(w) \int^w s^*(x)g(x)dx \\ \delta F(w) &= \delta G(w) + \lambda \bar{F}(w) \int^w s^*(x)dG(x)\end{aligned}$$

Solving this integral (2nd order Volterra) equation must be done numerically. However, I can re-arrange the above equation to get:

$$\frac{F(w) - G(w)}{\bar{F}(w)} = \frac{\lambda}{\delta} \int^w s^*(x)dG(x) > 0$$

Which shows that $G(w)$ (first-order) stochastically dominates $F(w)$.

Equating flow in with flow out of *employment at wage equal to* $w > w_R$, I get:

$$\begin{aligned}\lambda u s^* f(w) + \lambda(1 - u)f(w) \int^w s^*(x)g(x)dx &= [\delta + \lambda s^*(w)\bar{F}(w)][(1 - u)g(w)] \\ \delta f(w) + \lambda f(w) \int^w s^*(x)dG(x) &= [\delta + \lambda s^*(w)\bar{F}(w)]g(w) \\ \delta f(w) &= \delta g(w) + \lambda[\bar{F}(w)s^*(w) - f(w) \int^w s^*(x)dG(x)]\end{aligned}$$

Which, as expected, is just the derivative of the above 2nd order Volterra equation.

C.1.2 Firm's Problem

Next, we must derive an expression for the steady-state number of workers hired at a posted wage w , $l(w)$. Let there be N total workers in the economy, then:

$$l(w) = \lim_{\epsilon \rightarrow 0} N \left[\frac{G(w + \epsilon) - G(w)}{F(w + \epsilon) - F(w)} \right]$$

The firm must receive equal profit from all offered wages (or else it would not offer said wages in equilibrium). In other words, we must have:

$$\pi(w) = (p - w)l(w) = \bar{\pi} \quad \forall w$$

In other words, the steps to solve the general equilibrium model are as follows:

1. Given parameters, guess a wage distribution function
2. Solve for $s^*(x)$ — this involves solving an integral equation
3. Solve for $G(x)$ — this also involves solving an integral equation
4. Check if $\pi(w) = \bar{\pi} \quad \forall w$. Repeat.

C.1.3 Closed-Form Solution

Consider a simplification that provides an explicit solution. In particular, let us extract the important contribution of endogenous search effort, namely, that it decreases with wage.

To this end, assume that

$$s(w) = 1 - w$$

where I have normalized the support of w : $w \in [0, 1]$.

Then, the integral equation with $G(w)$ reduces to:

$$\begin{aligned}\delta F(w) &= \delta G(w) + \lambda \bar{F}(w) \left[\int^w dG(x) - \int^w x dG(x) \right] \\ &= \delta G(w) + \lambda \bar{F}(w) [G(w) - wG(w) + \mathbf{G}(w)] \\ &= G(w) [\delta + \lambda \bar{F}(w)(1 - w)] + \lambda \bar{F}(w) \mathbf{G}(w)\end{aligned}$$

Where I have used the fact that:

$$\begin{aligned}\int^w x dG(x) &= -w(1 - G(w)) + \int^w (1 - G(x)) dx \\ &= wG(w) - \mathbf{G}(w)\end{aligned}$$

where $\mathbf{G}(w)$ is the antiderivative of $G(w)$.

Re-arranging, I get:

$$\begin{aligned}\frac{\delta F(w)}{\lambda \bar{F}(w)} &= G(w) \frac{\delta + \lambda \bar{F}(w)(1 - w)}{\lambda \bar{F}(w)} + \mathbf{G}(w) \\ a(w) &= G(w)b(w) + \mathbf{G}(w) \\ a'(w) &= G'(w)b(w) + G(w)b'(w) + G(w) \\ \frac{a'(w)}{b(w)} &= G'(w) + G(w) \frac{1 + b'(w)}{b(w)} \\ q(w) &= y' + yp(w)\end{aligned}$$

Where I have renamed variables to simplify the final expression. Next, let $\mu(w) =$

$e^{\int p(w)dw}$. We can solve this ODE:

$$\mu(w)[y' + p(w)y] = \mu(w)q(w)$$

$$[\mu(w)y]' = \mu(w)q(w)$$

$$G(w) = \frac{\int \mu(w)q(w)dw + C}{\mu(w)}$$

Assuming $G(w)$ and $F(w)$ are both differentiable, I get:

$$\delta f(w) = g(w)[\delta + \lambda \bar{F}(w)(1 - w)] + G(w)\lambda[-f(w)(1 - w) - \bar{F}(w)]...$$

$$- \lambda f(w)\mathbf{G}(w) + \lambda \bar{F}(w)G(w)$$

$$f(w)[\delta + G(w)\lambda(1 - w) + \lambda \mathbf{G}(w)] = g(w)[\delta + \lambda \bar{F}(w)(1 - w)]$$

$$l(w) = \frac{\delta + G(w)\lambda(1 - w) + \lambda \mathbf{G}(w)}{\delta + \lambda \bar{F}(w)(1 - w)}$$

Notice, from our original equation, that:

$$\lambda \mathbf{G}(w) = \frac{\delta F(w) - G(w)[\delta + \lambda \bar{F}(w)(1 - w)]}{\bar{F}(w)}$$

Plugging into the expression for $l(w)$, I get:

$$\begin{aligned} l(w) &= \frac{\bar{F}(w)[\delta + G(w)\lambda(1 - w)] + \delta F(w) - G(w)[\delta + \lambda \bar{F}(w)(1 - w)]}{\bar{F}(w)[\delta + \lambda \bar{F}(w)(1 - w)]} \\ &= \frac{\bar{F}(w)\delta + \delta F(w) - G(w)\delta}{\bar{F}(w)[\delta + \lambda \bar{F}(w)(1 - w)]} \\ &= \frac{\delta \bar{G}(w)}{\bar{F}(w)[\delta + \lambda \bar{F}(w)(1 - w)]} \\ &= H(F(w)) \end{aligned}$$

Where I have plugged in the derived closed form solution for $G(w)$ above to get $l(w)$ as a function of $F(w)$ only. Then:

$$H(w)(p - w) = \bar{\pi}$$

$$H'(w)(p - w) - H(w) = 0$$

$$\frac{H'(w)}{H(w)} = (p - w)^{-1}$$

$$H(w) = e^{\frac{1}{2(p-w)^2}}$$

In other words, solving for $F(w)$ involves solving the fixed point problem above, for every w .

I leave the application of this general equilibrium model for future work.

C.2 Simulation

C.2.1 Parametrization

As in Christensen et al. [2005], I assume the cost function to be of the following form for simulation:

$$c(s) = c_0 \frac{s^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$$

Following Lentz and Mortensen [2010], I use the following parametrization:

$$c_0 = 0.10$$

$$\gamma = 5.00$$

$$\rho = 0.05$$

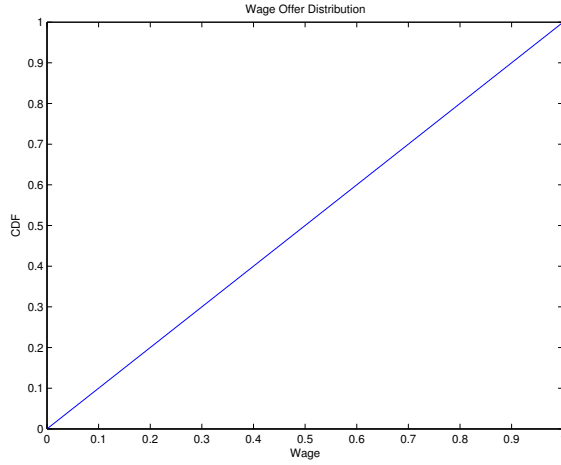
$$\delta = 0.25$$

$$\lambda = 0.125$$

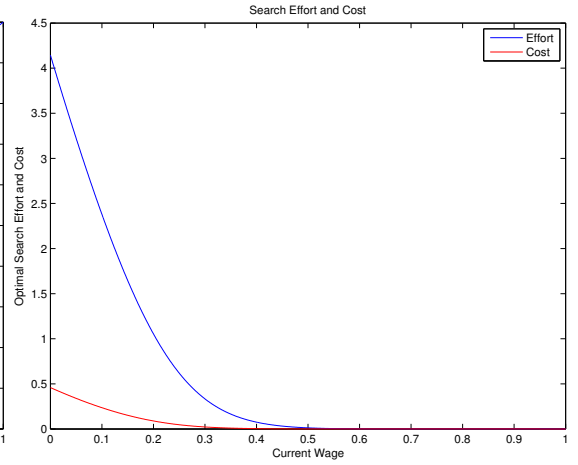
Using the above cost function, the search intensity equation (3.2.4) reduces to:

$$\begin{aligned} c'(s^*(w)) &= \lambda \int_w \frac{\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} \\ c_0 s^*(w)^{\frac{1}{\gamma}} &= \lambda \int_w \frac{\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} \\ s^*(w) &= \left(\frac{\lambda}{c_0} \int_w \frac{\bar{F}(x)dx}{\rho + \delta + \lambda s^*(x)\bar{F}(x)} \right)^\gamma \end{aligned}$$

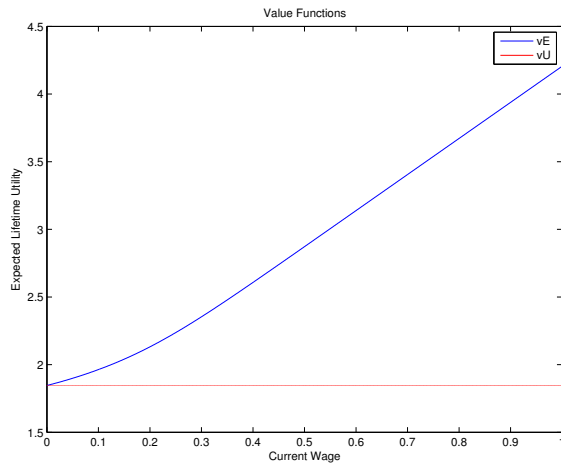
The above is a Volterra equation of the second kind. To solve it numerically, I solve it backwards using a grid of wage values, noting that at the highest wage offered \bar{w} , the integral “disappears”



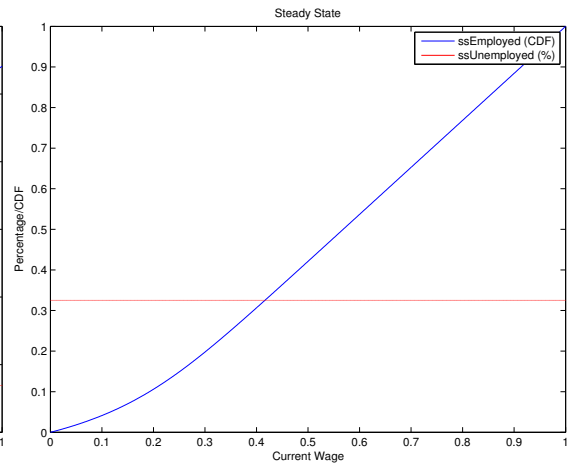
(a) (Exogenous) Wage Offer Distribution



(b) Search Effort and Cost



(a) Value Functions



(b) Steady State

C.2.2 Procedure

To simulate this continuous time model, I use the following procedure (note that for the exponential distribution, when I say $\sim \exp(y)$, it means drawing from the exponential distribution with mean $1/y$):

1. Everyone starts unemployed ($t = 1$)

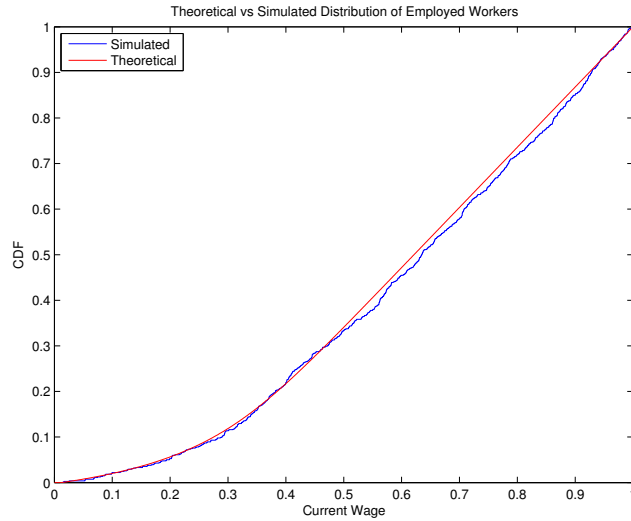
2. Draw unemployment duration before accepting job offer $\tau_u \sim \exp(\lambda s^* \bar{F}(w_R))$
3. Draw $w \sim F(w|w \geq w_R)$

Now everyone is employed ($t = 2$)

1. Draw $\tau_\delta \sim \exp(\delta)$
2. Draw $\tau_{j2j} \sim \exp(\lambda s^*(w) \bar{F}(w))$
3. Draw $w \sim F(w'|w' \geq w)$
4. Take $\min(\tau_\delta, \tau_{j2j})$

Do this many times, say, for $t = 1000$, and see distribution of wages and unemployment and employment rates.

Using the optimal tax system (subject to no debt), and the parameters described above, over 1000 “events” and 1000 individuals, I get the following comparison plot between simulated and theoretical distribution of wages of employed workers:



Theoretical steady-state unemployment is 18.27%, while simulated unemployment is 18.20%.

I try both uniform and lognormal exogenous wage offer distributions. For unbounded distributions like the lognormal, it is important to extrapolate properly after the largest grid point in wages, since as sample size increases, larger wages inevitably occur. One way to do this is to increase the grid size, a better way is to include a far off point and set the PDF equal to zero there, and interpolate between the resulting points. The latter is the approach used in this chapter.